

/instituut voor de
Nederlandse taal/

Lemmatiseerprincipes voor GiGaNT, het centrale lexicon van het INT

Uitgebreide en herziene versie van *Het lemma in het GiGaNT-lexicon* van Marijke Mooijaart (2012).

Katrien Depuydt, Thomas Haga, Marijke Mooijaart
m.m.v. Roland de Bonth, Griet Depoorter, Katrien Van pellicom en Boukje Verheij

Inhoudsopgave

Woord vooraf	5
1 Inleiding	6
1.1 Het GiGaNT-lexicon.....	6
1.2 Lemmata in GiGaNT	7
1.3 De functie van GiGaNT	8
1.4 Deze working paper	8
1.5 Verklaring van tekens en typografie	9
2 De vorm van het lemma	10
2.1 Canonieke vorm.....	10
2.2 Spellingvorm.....	10
2.3 Vormvariatie	10
2.4 De lemmagrenzen	11
2.5 Affixen in het lemma	12
2.6 Lemmatisering van voornaamwoorden en telwoorden	12
2.7 Vreemde woorden (leenwoorden met uitheemse kenmerken).....	12
2.8 Eigennamen	13
3 Het lemma en eventuele varianten	13
3.1 Gewone tegenover gemarkeerde varianten.....	13
3.2 Variatie in woordfamilies	14
3.3 Desambiguering.....	15
3.4 Behoud van de oorspronkelijke woordstructuur	15
3.5 Splitsing en samenvatting	16
3.6 Extra lemma's.....	17
4 De lemmagrenzen	18
4.1 Samenkoppelingen.....	18
4.2 Meerwoordsexpressies.....	19
4.3 Clitische combinaties	20
4.4 Afkortingen.....	20
4.5 Foutieve vormen	21
5 De weergave van affixen	21
5.1 Handhaving van oorspronkelijke affixen.....	21
5.2 Standaardvormen	22
6 Lemmatisering van voornaamwoorden en telwoorden	22
6.1 Persoonlijk voornaamwoord	23
6.2 Bezittelijk voornaamwoord.....	24
6.3 Aanwijzend voornaamwoord	24
6.4 Reflexief voornaamwoord	24
6.5 Reciproque voornaamwoord.....	24
6.6 Onbepaald voornaamwoord.....	24
6.7 Telwoorden	25
7 Leenwoorden	25
7.1 Aansluiting bij de oorspronkelijke vorm	25

7.2 Meervoudige ontleningen	26
7.3 Verouderde vreemde spellingen	26
8 Eigennamen.....	27
8.1 Relatie tot soortnamen	27
8.2 De werkwijze per categorie	27
8.2.1 Inheemse (Germaanse) voornamen	27
8.2.2 Uitheemse voornamen en andere persoonsnamen.....	28
8.2.3 Appellatieven als toenaam.....	28
8.2.4 Moderne namen.....	29
8.2.5 Inheemse toponiemen.....	29
8.2.6 Uitheemse toponiemen en toponiemen met uitheemse elementen	30
9 Detailproblemen en voorbeelden bij 3 t/m 5	30
Bij 3 Het lemma en eventuele varianten	30
9.3.1 Het al of niet volgen van regelmatige klankveranderingen	30
9.3.2 Enkele specifieke klankwijzigingen 1: apocope van -e.....	31
9.3.3 Enkele specifieke klankwijzigingen 2: syncope van -d-	32
9.3.4 Onregelmatige en niet goed traceerbare klank- en vormwijzigingen.....	33
9.3.5 Gewone tegenover gemarkeerde varianten in de moderne standaardvorm.....	33
9.3.6 Volksetymologische wijzigingen en woordspelingen	34
9.3.7 Morfologie: huidige regels worden gevolgd	34
9.3.8 Samenval en splitsing.....	35
Bij 4 De lemmagrenzen	36
9.4.1 Enkele specifieke verbindingen: naamwoorden met een genitief	36
9.4.2 Enkele specifieke typen: combinatie met een werkwoord	36
9.4.3 Enkele specifieke verbindingen: combinatie met een deelwoord	36
Bij 5.2 Affixen	37
9.5.2.1 Afwijkingen.....	37
9.5.2.2 Enkele specifieke suffixen 1: -e	37
9.5.2.3 Enkele specifieke suffixen 2: -(e)de/-te.....	38
9.5.2.4 Enkele specifieke suffixen 3: feminiserende suffixen.....	38
9.5.2.5 Enkele specifieke suffixen 4: diminutiefsuffixen	39
9.5.2.6 Leensuffixen en suffixen in leenwoorden	39
Bij 5.3 Vreemde woorden	40
9.5.3.1 Voorbeelden van aanpassingen	40
9.5.3.2 Verdwenen vreemde woorden.....	40
9.5.3.3 Voorbeelden van meervoudige ontlening.....	41
Literatuur.....	42

Woord vooraf

Dit document is een uitgebreide en herziene versie van Mooijaart, M. (2012). Het lemma in het GiGaNT-lexicon. *INL Working Papers - Taalbank Nederlands 4*. Leiden. Het beschrijft de principes voor toekenning van een modern Nederlands lemma aan diachroon Nederlands zoals die zijn ontwikkeld voor het GiGaNT-lexicon.

Het uitgangspunt van de vorige versie van de working paper was de historische lexiconcomponent (GiGaNT-Hilex). Sinds de verschijning van dat document is er niet alleen verder gewerkt aan de uitbouw van GiGaNT-Hilex, maar is er ook gewerkt aan de moderne lexiconcomponent (GiGaNT-Molex) en aan de integratie van deze twee lexiconcomponenten. Daarnaast zijn de lemmatiseerprincipes toegepast bij de ontwikkeling van training- en evaluatiesets¹ waarmee de taggers en lemmatiseerders in het GaLAHaD-platform voor de taalkundige verrijking van diachroon corpusmateriaal getraind en geëvalueerd zijn.² De lemmatiseerprincipes die destijds geformuleerd waren, zijn daardoor uitvoerig toegepast en getoetst. Dit heeft aanleiding gegeven tot verdere aanscherping en hier en daar tot andere keuzes.

Het resultaat daarvan is te vinden in deze nieuwe working paper. Voor de tagset die gebruikt is voor GiGaNT en voor de training- en evaluatiesets, verwijzen we naar Haga T., Depuydt, K. & Does, J. de et. al. (2024). *Tagset voor Diachroon corpusmateriaal van het Nederlands (TDN)*. INT working papers 1 (Versie 2.0) (Instituut voor de Nederlandse Taal).

Katrien Depuydt
Leiden, november 2024

¹ Zie <https://github.com/INL/galahad-corpus-data>.

² Zie <https://portal.clarin.ivdnt.org/galahad/>.

1 Inleiding

1.1 Het GiGaNT-lexicon

In GiGaNT (Groot Geïntegreerd Lexicon van de Nederlandse Taal), het diachroon morfosyntactisch computationele lexicon van het Nederlands dat door het Instituut voor de Nederlandse Taal (INT) wordt ontwikkeld, worden alle woorden van het Nederlands, vanaf de zesde eeuw tot nu, opgeslagen.³ Het lexicon vormt het fundament van de kennisbank van het INT⁴ en koppelt alle lemmata en hun vormvarianten uit de hedendaagse en historische lexicale databanken van het INT aan één uniek en persistent superlemma-ID. Hiermee wordt informatie gecentraliseerd over woorden en hun mogelijke spellingen door de tijd heen met links naar gedateerde bewijspplaatsen (attestaties), het biedt informatie over het paradigma, bevat morfologische informatie en in beperkte mate ook betekenisinformatie.⁵ De link met de gekoppelde lexicale databanken wordt ook behouden door het overnemen van de persistente ID's van de informatie die vanuit die databanken in GiGaNT is opgenomen.

GiGaNT wordt modulair ontwikkeld. Er is een historische lexiconcomponent (GiGaNT-Hilex), die primair gebaseerd is op de historische woordenboeken van het Nederlands⁶ en op termijn aangevuld zal worden vanuit nieuw diachroon corpusmateriaal. Daarnaast is er een moderne lexiconcomponent (GiGaNT-Molex), die wordt gevoed door hedendaags Nederlands corpusmateriaal en de hedendaags Nederlandse lexica van het INT. De beide componenten worden gekoppeld en komen samen in het superlemma.⁷

Ieder lemma en door de koppeling ook ieder superlemma in het GiGaNT-lexicon is voorzien van de correcte Nederlandse lemmavorm en van een woordsoort. Dit gebeurt op etymologische gronden. Superlemmata worden van elkaar onderscheiden op basis van de etymologie van een woord. Etymologie is dus leidend bij de koppelingen van de lexicale databanken aan GiGaNT.

Voor de woordsoortinformatie in GiGaNT en in de door het INT ontwikkelde lexicale producten die aan GiGaNT zijn gekoppeld wordt gebruik gemaakt van de TDN (Tagset voor Diachroon corpusmateriaal van het Nederlands).⁸

Het toekennen van hedendaags Nederlandse lemmavormen in GiGaNT en in de gekoppelde INT-lexica gebeurt ook volgens zorgvuldig gedefinieerde principes, die in deze working paper verder beschreven worden. Deze principes worden ook gebruikt bij het

³ Voor een beschrijving van GiGaNT zie Ruitenbergh, De Does en Depuydt (2010).

⁴ Zie Depuydt, Tiberius & Heylen (te verschijnen).

⁵ Betekenisinformatie zit in de huidige versie van het lexicon in de vorm van glossen ter onderscheiding van homoniemen in de moderne lexiconcomponent van GiGaNT, en via de link vanuit de bewijspplaatsen (attestaties) naar de historische woordenboeken in de historische lexiconcomponent.

⁶ Dat zijn het *Woordenboek der Nederlandsche Taal* (WNT), het *Middelnederlandsch Woordenboek* (MNW), het *Vroegmiddelnederlands Woordenboek* (VMNW) en het *Oudnederlands Woordenboek* (ONW), te raadplegen via <https://gtb.ivdnt.org/search/>. MNW en WNT zijn reeds opgenomen in GiGaNT. VMNW en ONW moeten nog worden verwerkt.

⁷ Aan deze koppeling wordt bij het verschijnen van deze working paper nog steeds gewerkt.

⁸ Haga, Depuydt en De Does m.m.v. De Bonth en Geirmaert (2024). Deze tagset is onder andere gebaseerd op de eertijds voor GiGaNT ontwikkelde tagset (Ruitenbergh en Van pellicom (2012)) en wordt ook de standaard voor de taalkundige verrijking van het corpusmateriaal van het INT.

lemmatiseren van diachroon corpusmateriaal, dat een essentieel onderdeel vormt van de academische lexicografische beschrijving van het Nederlands. In geval er bij de corpuslemmatisering afgeweken wordt van deze principes, zal dit expliciet vermeld worden.

1.2 Lemmata in GiGaNT

GiGaNT bestaat uit lemmata, die voorzien zijn van een woordsoort en gelinkt met daarmee corresponderende woordvormen. In GiGaNT-Molex zijn die woordvormen onderdeel van het paradigma, in GiGaNT-Hilex zijn die woordvormen gekoppeld aan attestaties (vindplaatsen). Op termijn moeten ook attestaties gekoppeld worden aan de woordvormen in Molex en zal het paradigma, dat nu alleen hedendaags Nederlandse woordvormen bevat, moeten aangevuld worden met diachrone data.

De lemmata in GiGaNT zijn voorzien van een correcte lemmavorm die vooraf vastgestelde, controleerbare principes volgt. Dit bevordert de voorspelbaarheid en doorzichtigheid van het lemmasysteem voor bewerkers en gebruikers. Zo kan bovendien ook nieuw materiaal, bijvoorbeeld uit corpora, op dezelfde manier worden gelemmatiseerd en in GiGaNT worden opgeslagen.

Het uitgangspunt is een vaste identiteit voor een woord gedurende de hele taalgeschiedenis, en deze identiteit wordt niet alleen zichtbaar omdat het woord een persistent superlemma-id heeft, maar wordt ook gereflecteerd in de vorm van het lemma.

Voorbeeld: lemma WERELD

- Ecco sia sundiga in thionda an **uueroldi** [ID 133342 WERELD noun dat.sg] hatton ricduom. Zie, zij zondaars en zij die gedijen in de wereld, hadden rijkdom. WPs (hs. I) 072,12 Zuid-Nederrijn, Nederrijn, 901-1000 (kopie 1701-1800)
- Dye ... gebenedide droppelkens, die wij hier in deser aerden der **warelt** [ID 133342 WERELD noun gen.sg.] daer van mogen vercrigen, S. Bern. S. 1, 108c., Noord-Oost Nederland, 1484-1485
- In t'iaer des **vveerelts** [ID133342 WERELD noun gen.sg.] vier duust ongezvvicht En zessenveertich, vvert dees machtighe stad ghesticht, v. VAERNEWIJCK, *Vl. Audtvr.* B iijij r^o [1562].
- De knegt ... kwam zeggen, dat Dominé Q. met een Ouderling daar waren ... Gerrit presenteerde Caap-wyn en pypen. Want zie, Santje, ik heb tegen de menschen niets ter **waereld** [ID 133342 WERELD noun dat sg.], WOLFF en DEKEN, *Leev.* 1, 89 [1784].

In het lexicon kennen we aan de verschillende spelling- en vormvarianten van een woord, ontstaan door dialectische en chronologische ontwikkelingen en door verschillende spellingconventies, één moderne lemmavorm toe. Zo is het voor de gebruiker gemakkelijker om alle type varianten van dat woord te verzamelen. In principe wordt daarom één lemmavorm toegekend.

Daarvan wordt afgeweken als een woord in het hedendaags Nederlands ook geaccepteerde lemmavarianten heeft (bijvoorbeeld TORNOOI naast TOERNOOI) Een enkele

lemmavorm volstaat ook niet wanneer een woord zich in de loop van de geschiedenis splitst (bijvoorbeeld in het geval van KRUID en KRUIT) of wanneer verschillende woorden samenvallen in de loop der tijd (zoals SLIJK en SLIK). Zie verder 3.6.

De hoofdvragen in deze nota zijn:

- a. welke vorm krijgt een lemma;
- b. volgens welke criteria worden bepaalde vormen en varianten onder één bepaald lemma ondergebracht.

1.3 De functie van GiGaNT

Het GiGaNT-lexicon functioneert als tool (of reeks tools) én als thesaurus (zie ook Ruitenbergh en Van pellicom 2012: 6-7). De toolfunctie houdt wat het lemma betreft onder andere de volgende mogelijkheden in:

- zoeken in historisch tekstmateriaal;
- verrijking van historisch en modern tekstmateriaal met standaardlemmavormen en daaraan gekoppelde informatie;
- ondersteuning van OCR bij productie en correctie;
- opsporen van tot nu toe onbeschreven woorden, zowel oude als recente (datamining en datasuppletie);
- koppeling van de online woordenboeken van het INT aan elkaar en andere woordenboeken;
- koppeling van tekstmateriaal aan de online woordenboeken van het INT;
- centralisering van de morfosyntactische beschrijving van de woordenschat.

Als thesaurus is GiGaNT een databank van velerlei gegevens over de Nederlandse woordenschat en de verbanden daartussen. De gebruiker kan die vanuit elke informatiecategorie die hij bevat, benaderen en desgewenst inzicht krijgen in alle relaties die erin opgesloten liggen. Zo kan het invoeren van een woordvorm uit een tekst, of van een bekende oudere vorm leiden naar het lemma, en dat weer naar de informatie in de historische woordenboeken.

Zowel voor diachroon als synchroon onderzoek biedt GiGaNT het benodigde materiaal. Hoe dit alles technisch wordt gerealiseerd, is niet het onderwerp van deze nota; zie hiervoor Ruitenbergh, De Does en Depuydt (2010).

1.4 Deze working paper

In dit stuk gaan wij in op de op de vorm van de lemma's (2) en op de selectie van de daaronder vallende varianten. Deze varianten worden gevonden onder diverse soorten klank- en vormvarianten (3) en in variatie in structuur, samengevat in de term lemmagrenzen (4).

Daarna volgen besprekingen van enkele bijzondere lemmatypen en probleemcategorieën, namelijk affixen (5), voornaamwoorden (6), vreemde woorden (7) en eigennamen (8). Hoofdstuk 9 bestaat uit verdere illustraties en verdere uitwerkingen van de voorgaande hoofdstukken.

1.5 Verklaring van tekens en typografie

Voorbeelden

Voorbeelden (linkerkolom) komen meestal uit de vier historische woordenboeken van het INT, maar ook uit ander historisch en modern tekstmateriaal. Deze woordvormen kunnen ter wille van de duidelijkheid soms gestandaardiseerd zijn. Hierbij staat het lemma altijd in klein kapitaal (rechterkolom).

Aanduidingen van meerledige lemma's in deze nota:

- **komma** (,) scheidt de lemma's (klein kapitaal) van lemmavarianten (cursief klein kapitaal);
- **plus** (+) wordt gebruikt bij de analyse van een token dat uit meer woorden bestaat;
- **slash** (/) betekent: afzonderlijke lemma's voor afzonderlijke woorden;
- **sluisteken** (|) duidt op ambiguïteit: meerdere verschillende lemma's zijn mogelijk voor één woordvorm;
- **spatie** verbindt elementen van een meerwoordsexpressie.

2 De vorm van het lemma

2.1 Canonieke vorm

De lemmavorm is altijd de canonieke vorm, dus de vorm zonder flexie (bijvoorbeeld bij zelfstandig naamwoord enkelvoud, nominatief, bij werkwoord infinitief), behalve wanneer de geflecteerde vorm gelexicaliseerd is (GEBOREN, SLEETS, TROPEN). Als er sprake is van suppletieve vormen, zoals bij de trappen van vergelijking van sommige adjectiva (*beter, best* bij GOED; *minder, minst* bij WEINIG), dan krijgen deze suppletieve vormen een eigen lemma. Het werkwoord *zijn* vormt hierop een uitzondering: *ben, is, was, waren* en *gewezen* krijgen alle het lemma ZIJN.

2.2 Spellingvorm

De spelling volgt de meest recente officiële spellingregels, op dit moment die van 2016, zoals vastgelegd in de [Technische handleiding spelling](#) van de Nederlandse Taalunie. Bij verouderde woorden die niet volgens de algemene regels herspeld kunnen worden en niet in hedendaagse woordenboeken te vinden zijn maar wel in het WNT, geldt in principe de hoofdvorm van het WNT, aangepast aan de recentste spellingregels. De gedachte hierachter is dat het WNT de standaardvorm heeft vastgesteld van regionaal gevarieerde woordvormen. Spellingen van vreemde woorden komen bij de vormvariatie aan de orde.

2.3 Vormvariatie

(zie verder hoofdstuk 3)

Naast spellingvariatie (verschillende weergaven van eenzelfde klank, zoals in *ik - ick*) is er vormvariatie: vormverschillen binnen eenzelfde woord. Ze kunnen betrekking hebben op de klanken van de morfemen, op verbindingsklanken in gelede woorden en op uitgangen, zoals in resp. *vogel - veugel, beminlijk - beminnelijk* en *luid - luide*. In principe kiezen we voor het lemma de meest gebruikelijke moderne vorm, als het woord nog bestaat. Hierbij toetsen we ons eigen oordeel aan frequentiegegevens uit het [Corpus Hedendaags Nederlands](#) en aan al of niet expliciete oordelen in de historische woordenboeken en externe lexicale bronnen. Wanneer er naast de meest gebruikelijke moderne vorm nog een variante vorm bestaat in het hedendaags Nederlands, dan wordt deze in het lexicon wel als variant opgenomen.

Bij nog bestaande woorden

Het lemma is de meest gebruikelijke, moderne vorm van het woord in het Standaardnederlands:

mnl. huus	HUIS
vnnl. moêr	MOEDER
mnl. sterre	STER
mnl. bernnen	BRANDEN

Soms wijkt de lemmavorm in het simplex volgens bovengenoemd principe af van de lemmavorm van datzelfde element in een samengesteld woord. Verbindingsklanken en uitgangen volgen de moderne vorm.

mnl. arg	ERG
mnl., nnl. arglist	ARGLIST
nnl. hemelblauw	HEMELSBLAUW

Bij verdwenen woorden

Het lemma is de etymologische vorm met verwerking van regelmatige klankwijzigingen, gespeld volgens de nu geldende regels.

onl. mūrhuodil ‘muurwacht’	MUURHOEDEL
mnl. saen ‘spoedig’	ZAAN
mnl. stiden ‘stevig worden’	STIJDEN
mnl. vergierer ‘wijnroeier’	VERGIERDER

Bij woorden die zich splitsen of samenvallen in de loop van de geschiedenis

Variatie kan leiden tot het ontstaan van verschillende nieuwe woorden uit eenzelfde woord (splitsing), omgekeerd kunnen verschillende woorden samenvallen tot één woord (samenval). Er worden in het eerste geval aan de vormen van vóór de splitsing twee lemma’s toegekend in het lexicon. In de corpuslemmatisering wordt rekening gehouden met de betekenis.

Splitsing:

mnl. cruut	KRUID KRUIT
------------	---------------

Samenval:

mnl. brusen	BRUISEN ‘schuimen, gisten’
mnl. brusschen	BRUISEN ‘bruisen, schuimen; stormen’

2.4 De lemmagrenzen

(zie verder hoofdstuk 4)

Meerwoordsexpressies die bestaan uit ontleende verbindingen worden in het lexicon als geheel opgenomen en in corpora als geheel gelemmatiseerd.

mnl. kyrie eleison	KYRIE ELEISON
nnl. toute à toi	TOUTE À TOI

Inheemse verbindingen daarentegen worden in corpora altijd afzonderlijk gelemmatiseerd, maar kunnen in het lexicon wel als ingang worden aangetroffen. Dat laatste is bijvoorbeeld het geval bij telwoorden of omdat daar in een gekoppelde lexicale bron aanleiding toe is.

mnl. o wach	O WACH (lemma in het MNW)
nnl. acht miljoen	ACHT MILJOEN (telwoord)
nnl. achterwege laten	ACHTERWEGE LATEN (Molex)
nnl. heg noch steg	HEG / NOCH / STEG
nnl. van kindsbeen af	VAN / KINDSBEEN / AF

Meerwoordsexpressies die opgenomen zijn in het lexicon worden gelinkt aan hun samenstellende delen.

Bij clitische verbindingen en samenkoppelingen met Nederlandse (niet hetzelfde als inheemse) elementen worden de moderne woordgrenzen in acht genomen:

mnl. sconinxs	DE + KONING
mnl. tnacht	TE + NACHT
vnnl. lanckx henen	LANGSHEEN
vnnl. wel bedocht	WELBEDACHT

2.5 Affixen in het lemma

(zie verder hoofdstuk 5)

Affixen worden altijd in hun eigen vorm in het lemma weergegeven, dus niet weggelaten of vervangen:

onl. gihōrsamighēt	GEHOORZAMIGHEID (niet: GEHOORZAAMHEID)
mnl. hackelen	HAKKELEN (niet: HAKKEN)
mnl. broekacht	BROEKACHT (niet: BROEKIG)

2.6 Lemmatisering van voornaamwoorden en telwoorden

(zie verder hoofdstuk 6)

Bij voornaamwoorden geldt de nominatief (subjectvorm) als lemmavorm. De objectvormen krijgen ook de subjectvorm als lemma.

mnl. du, di, dijns	DU
--------------------	----

Hoofdtelwoorden krijgen hun eigen lemma. Ze worden voluit geschreven bij getallen van nul tot en met honderd en bij honderdtallen en duizendtallen.

honderd	HONDERD
tienduizend en een	10001
mnl. vierwarf twintich	VIERWERFTWINTIG

Rangtelwoorden krijgen ook hun eigen lemma, waarbij het onderscheid tussen de suffixen *-de* en *-ste* behouden blijft.

tweede	TWEEDE
mnl. tweeste	TWEESTE

2.7 Vreemde woorden (leenwoorden met uitheemse kenmerken)

(zie verder hoofdstuk 7)

Deze krijgen evenals erfwoorden de modern-Nederlandse vorm:

nnl. vanielje	VANILLE
vnnl. voyaedse	VOYAGE
vnnl. affinagie	AFFINAGE

Voor verdwenen vreemde woorden gelden in principe dezelfde regels als voor inheemse woorden, maar omdat de klankontwikkelingen van vreemde woorden binnen het Nederlands veel onregelmatiger zijn, is hier wel eens een ad-hocregel toegepast. Zo krijgen *battement* en

esbattement elk een eigen lemma als frequent voorkomende varianten, hoewel *es-* geen prefix is.

2.8 Eigennamen

(zie verder hoofdstuk 8)

We beschouwen persoonsnamen, toponiemen en namen van organisaties, instellingen en dergelijke als eigennamen. Ook gepersonificeerde voorstellingen van begrippen, voorwerpen en dergelijke, die bijvoorbeeld voorkomen in (blij)spelen (zoals *Beroemen*, *Gratie*, *Obstinatie* en *Reden*) worden als eigenaam beschouwd.

Eigennamen worden altijd als naam gelemmatiseerd. Bij historische namen (tot 1500) kan dat lemma soms samenvallen met het appellatief waarop de naam gebaseerd is. In andere gevallen worden namen in beperkte mate gestandaardiseerd in het lemma, afhankelijk van het type naam. Familienamen worden niet gestandaardiseerd, tenzij er sprake is van verschillende spellingen van een naam die duidelijk naar dezelfde referent verwijst.

Voornamen, familienamen en plaatsnamen kunnen uit meerdere woorden bestaan. In al deze gevallen wordt er één lemma toegekend. Bij namen van organisaties, namen van boeken, films, toneelstukken en dergelijke die uit meerdere woorden bestaan, krijgt elk woord een eigen lemma.

Piet Hein	PIET HEIN
Jan-Willem	JAN-WILLEM
Amstelre damme	AMSTERDAM
Depuydt	DEPUYDT
Van den Berghe	VAN DEN BERGHE
New York	NEW YORK
Pluk van de Petteflet	PLUK / VAN / DE / PETTEFLET

3 Het lemma en eventuele varianten

Het lemma functioneert als koepel voor allerlei diachrone en synchrone varianten. Hieronder worden de verschillende typen varianten besproken. Daarbij komen tegelijk de criteria aan de orde die bepalen welke van verschillende varianten het standaardlemma wordt.

3.1 Gewone tegenover gemarkeerde varianten

De vorm die in het moderne Nederlands het meest gebruikelijk is, geldt als lemma. Hieronder kunnen vormen vallen die eveneens tot de standaardtaal behoren, maar gemarkeerd zijn in hun gebruik. Wanneer van een lemma in het hedendaags Nederlands ook nog een vormvariant voorkomt, dan wordt deze ook in het lexicon opgenomen.

rouw, ruw	RUW, <i>ROUW</i>
einde, eind, end	EINDE, <i>EIND</i> , <i>END</i>
honig, honing	HONING, <i>HONIG</i>
vrouw, vrouwe	VROUW, <i>VROUWE</i>

Nog enkele specifieke gevallen:

Voorbeeld 1: *peuluw, poluw, peluw*: hiervan is *peluw* het meest gebruikelijk en komt nog voor in het hedendaags Nederlands: lemma wordt PELUW, ook in samengestelde woorden.

Voorbeeld 2: bij de *ie-uu*-variatie is per woord voor de moderne *ui*-vorm als lemma gekozen. Indien beide vormen nog bestaan, wordt de *ie*-vorm als variant opgenomen.

mnl., vnnl. bedieden	BEDUIDEN
mod.nl. beduiden	BEDUIDEN
rieken, ruiken	RUIKEN, <i>RIEKEN</i>

Zo ook STUREN en BESTUUR en de vele andere samengestelde verwante woorden. Bij *lui* en *lieden* en vele samengestelde woorden geldt als standaardvorm echter de *ie*-vorm (met uitzonderingen ARMELUI, JONGELUI, OUWELUI en RIJKELUI), waarbij in het lexicon ook de *ui*-variant (respectievelijk de *lieden*-variant) wordt opgenomen.

Reeksen klank- en spellingvarianten van één woord kunnen tot enkele groepen varianten worden teruggebracht.

Voorbeeld 3, *tegen*: spellingen met *t-*, *s-* en *tj-*: TEGEN, spellingen met *j-*: JEGEN.

mnl. tejeghens	TEGENS
mnl. jegenworde	JEGENWOORD

Voorbeeld 4: Veel vormvariatie komt voor bij dialectwoorden. De varianten *darie, derrie, daring, derring, darg, derg* ‘veenachtige grond’ (en daar weer spellingvarianten van) worden onder twee lemma’s ondergebracht: DERRIE voor de eerste twee, DERRING voor de overige. Klankwisselingen in de eerste syllabe worden dus genegeerd (*-er-ar*-variatie voor *-r* komt veel voor); bij *-ing* kan men aan een suffix denken (*-g* is daar weer een variant van).

Ook bij tussenklanken en buigingsuitgangen hanteren we de meest gehanteerde moderne vorm, en als die niet bekend is, de vorm van vergelijkbare samenstellingen. In sommige gevallen kan het betekenen dat in het lexicon ook een lemmavariant wordt opgenomen.

Voorbeelden tussenklanken:

hoerhuis	HOERENHUIS
hemelblauw	HEMELSBLAUW
straatplan	STRATENPLAN, <i>STRAATPLAN</i>

Is er geen duidelijke analogie, dan blijft de oude tussenklank gehandhaafd. De tussenklank *e* wordt hierbij volgens de moderne spellingregels aangepast. Het adverbiale suffix *-s* daarentegen rechtvaardigt een eigen lemma.

Voorbeelden uitgangen:

mnl. sint	SINT
nl. sinds	SINDS
mnl. elder	ELDER
nl. elders	ELDERS

3.2 Variatie in woordfamilies

De vorm van een modern lemma dat toegekend is aan een simplex, hoeft niet noodzakelijk te gelden voor de of alle daarmee samengestelde of afgeleide woorden. Dit geldt zowel voor de

historische als voor de moderne woordenschat. Zo zijn alle historische varianten van *branden*, o.a. *brennen*, *bernen*, *barnen*, *bornen* in GiGaNT als BRANDEN gelemmatiseerd, maar de vorm *barn-* is wel bewaard gebleven in BARNSTEEN. Het is lastig precies aan te geven wanneer in samengestelde woorden geüniformeerd wordt en wanneer ze hun verschillende vormen kunnen behouden. Zie de voorbeelden.

Voorbeeld 1: *lijtteken*, *lycteecken*, *licteken*, *lidteeken* met als lemma LITTEKEN. Toelichting: *-teken* conform de lemmatisering van het simplex, maar niet *lijk-* conform het simplex LIJK, omdat de moderne geassimileerde vorm van de samenstelling LITTEKEN is.

Voorbeeld 2:

mnl. bodelinge ‘ingewanden’	BEULING
nml. beuling ‘ingewanden’	BEULING
Hiernaast vmnl. bodel ‘darm’	BODEL (in VMNW geen <i>bodeling</i>)
NB: Uiteraard worden zuivere spellingaanpassingen wel altijd doorgevoerd in samengestelde woorden:	
<i>visc</i> , <i>vischbestand</i> , <i>vischfabriek</i>	VIS / VISBESTAND / VISFABRIEK

3.3 Desambiguering

Daar waar homoniemen in tekstmateriaal elk dezelfde variëte vormen kennen, wordt dit in het GiGaNT-lexicon zoveel mogelijk opgelost door verschillende lemma’s toe te kennen. Voorbeeld 1: *gard*, *garde* ‘roede’, ‘stok’, ‘puntig voorwerp’, naast *gaard*, *gaarde* ‘tuin’. Beide woorden komen onder andere voor als *gard*, *gaard*, *gaerd*, *garde*, *gaerde*. Daarnaast is er nog *gard*, *garde* ‘wachter, wacht’.

Bij lemmatisering is nu strikt onderscheiden in GARD ‘roede’, GAARD ‘tuin’ en GARDE ‘wacht’, ook in de samenstellingen waarvoor geen specifieke moderne Nederlandse vorm geldt. Is die er wel, dan wordt dat de lemmavorm, zoals bij DIERGAARDE (niet: DIERGAARD):

mnl. garde ‘stok; gaarde’ (maat)	GARD
nml. gard, garde ‘roede’	GARD
mnl. gaert, garde ‘tuin’	GAARD
nml. gaarde ‘tuin’	GAARD
nml. garde ‘lijfwatch’	GARDE

Voorbeeld 2: *-SEX-* is het moderne lemma voor de leenwoordelementen met de betekenis ‘zes’; tegenover *-SEKS-* voor alle woorden/woorddelen met de betekenis ‘geslacht’, met uitzondering van de Engelse leenwoorden (zoals *SEXY* en *SEXAPPEAL*), die een *x*-spelling kennen. Dit is conform de spellingvoorschriften.

Blijven ze als homoniem lemma gehandhaafd, zoals *VORST* ‘koning’ naast *VORST* ‘vriesskou’, dan worden ze in GiGaNT door een verschillend superlemma-ID onderscheiden, niet door de voor woordenboeken gebruikelijke homoniemnummers.

3.4 Behoud van de oorspronkelijke woordstructuur

Wanneer een woord in de loop van de tijd in het taalsysteem vervangen wordt door een ander dat daarop lijkt, wordt dat vervangende woord niet aangeduid met het lemma van de voorganger.

Voorbeeld 1: lidwoord onl. *that*, is etymologisch een ander woord dan het moderne lidwoord *het* (oorspronkelijk alleen pers.vnw.). Onl. lidwoord *that* krijgt dus lemma DAT en niet HET. Onl. *thie* en mnl. *die* (als lidwoorden) vallen beide wel onder DE, aangezien *de* zich heeft ontwikkeld uit de volle vorm.

Als de samenstellende delen in een afleiding verschillend zijn, dan kan dat eveneens voor verschillende woorden zorgen.

Voorbeeld 2: *edellike* wordt EDELLIJK, niet ADELLIJK, de basis is het adjectief EDEL en niet het zelfstandig naamwoord ADEL.

Ook kan verschil in morfologische structuur een rol spelen:

Voorbeeld 3: *weger* en *wager*, beide ‘persoon die weegt’, zijn eveneens verschillende woorden en geen vormvarianten. Het eerste is afgeleid van het werkwoord *wegen*, het tweede van het zelfstandig naamwoord *waag*, dat een ablautvorm is naast *wegen*. Het zijn dus twee lemma’s, WEGER en WAGER.

Voorbeeld 4: In het verleden werden in diverse functies *d*-vormen gebruikt terwijl nu *w*-vormen gewoon zijn: (de plaats) *daer hi sat* ‘waar hij zat’, (hij vertrok) *dat zij betreurde*. Hier worden de oorspronkelijke vormen gelemmatiseerd, aangezien het om woordvervanging gaat en niet om klankwisseling (Van der Horst 2008).

3.5 Splitsing en samenvall

Variatie kan leiden tot het ontstaan van nieuwe woorden uit eenzelfde woord (splitsing), omgekeerd kunnen verschillende woorden samenvallen tot één woord (samenvall).

Splitsing

Als van een woord een oude vorm bewaard is naast een nieuwe die hieruit ontstaan is, met in het moderne Nederlands ieder een eigen betekenis, is er in de jongste taalfase sprake van twee verschillende woorden met elk een eigen lemma. Die twee lemma’s moeten ook op het oorspronkelijke woord toegepast worden.

Voorbeeld 1: Een bekend geval van splitsing is die van onl. *krūt* in nnl. *kruid* ‘plant, specerij’ en *kruit* ‘ontploffbare poeder’. Het eerste woord bevat (onderliggend) de oorspronkelijke eindklank, een stemhebbende dentaal. De vorm met auslautverscherping in het enkelvoud, weergegeven in de officiële spelling, ging voor de stofnaam gelden. In dit geval zijn KRUID en KRUIT samen lemma tot het Nieuwnederlands; als er duidelijke betekenisdifferentiatie is worden afzonderlijke lemma’s toegekend.

Voorbeeld 2: Een ander bekend geval is de splitsing van mnl. *redene* in nnl. *rede* ‘denkvermogen; het spreken’ naast nnl. *reden* ‘oorzaak, grond, motief’.

mnl. redene (2 betekenissen)	REDE, REDEN
nl. reden ‘oorzaak, grond, motief’	REDEN
nl. rede ‘denkvermogen; het spreken’	REDE

Samenvall

In een enkel geval zijn in een modern woord twee verschillende historische woorden samengevallen. In veel gevallen, zoals bij adjectieven/bijwoorden op *-lijk* en

bijwoorden/voorzetsels zoals *in*, waarvan het bijwoord in het Middelnederlands een *-e*-uitgang had, geeft de woordsoort het verschil aan. In het lemma wordt geen onderscheid gemaakt.

Ook in andere gevallen heeft de samenvallende meestal in een vroege fase plaatsgevonden. Het hangt van het woord af, of, en voor welke fase dan twee lemma's worden toegekend.

Voorbeeld 1: In het Middelnederlands bestonden er twee werkwoorden *brusscen* en *brusen* met een bijna synonieme betekenis. De *-sch-* en *-s-* vielen in een late fase van het Middelnederlands samen in de inlaut en auslaut. In het Vroegnieuwlands bestond er daardoor geen vormelijk onderscheid tussen beide werkwoorden, waardoor deze zijn samengevallen.

mnl. brusen	BRUISEN 'schuimen, gisten'
mnl. brusschen	BRUISEN 'schuimen; stormen'

Voorbeeld 2:

mnl. *verblinden* (intransitief), *verblenden* (transitief) krijgen beide als lemma VERBLINDEN. Deze werkwoorden zijn in het Nieuwnederlands samengevallen, maar al in het Middelnederlands liepen de vormen door elkaar.

3.6 Extra lemma's

In principe hanteren we voor elke woordvorm maar één lemma. Meerdere lemma's kunnen worden toegekend in de volgende gevallen, deels hierboven besproken.

- 1) Bij splitsing en samenvallende (zie 3.5). Dit geldt alleen voor het lexicon.
- 2) Wanneer een woordvorm, meestal door een bepaalde typografie, twee woorden representeert:

plant(je)	PLANT + PLANTJE
(inter)nationaal	INTERNATIONAAL + NATIONAAL

- 3) Bij ambiguïteit: het valt, bijvoorbeeld door gebrek aan context, niet te bepalen of lemma 1 of lemma 2 van toepassing is:

vmnl. tijt (tegenwoordige tijd)	TIJDEN TIJGEN
nnl. enz. (lemma met of zonder -s)	ENZOVOORT ENZOVOORTS

- 4) Bij clitische verbindingen en andere gevallen van aaneengeschreven woorden die gesplitst moeten worden. Er worden aan één woordvorm twee lemma's toegekend, en wel in de volgorde van de elementen. Zie ook 4.3.

mnl. slaepstu (< slaeps du)	SLAPEN + DU
mnl. entie (< ende die)	EN + DE

- 5) Moderne afkortingen van verbindingen die grafisch een eenheid zijn, worden als verbinding gelemmatiseerd (vgl. clitics); elk woord krijgt een eigen lemma. Bij het lemmatiseren wordt de afkorting opgelost.

mod.nl. o.i.v.	ONDER + INVLOED + VAN
mod.nl. z.s.m.	ZO + SPOEDIG + MOGELIJK

Zo ook meer incidentele afkortingen als p.p., p.w., i.h. (midden), en za./zo.

- 6) Wanneer in het modern Nederlands min of meer gelijkwaardige varianten van een lemma voorkomen, dan wordt daar in de lemmatisering rekening mee gehouden. In het lexicon worden beide varianten opgenomen, voor de corpuslemmativering wordt aan één lemma de voorkeur

gegeven. Bij varianten met een lange en een kortere variant is dit op een enkele uitzondering na de lange variant.

weer, weder	WEDER, <i>WEER</i>
broeder, broer	BROEDER, <i>BROER</i>
kribbe	KRIBBE, <i>KRIB</i>
leuze	LEUZE, <i>LEUS</i>
hegge, heg	HEG, <i>HEGGE</i>

Zo ook bij *leder-leer*, *ledig-leeg* en *voederen-voeren* ‘eten geven’. Bij samengestelde woorden wordt de lemmavariant niet systematisch opgenomen en wordt per samenstelling voor de kortere variant (zoals *LEERBEWERKING*) dan wel de langere variant (zoals *LEDERSCHILDPAD*) gekozen.

4 De lemmagrenzen

Lemma's bestaan doorgaans uit één woord, maar kunnen ook bestaan uit meerdere woorden. De grammaticale context en de woordbetekenis bepalen of we met een verbinding of samenkoppeling te maken hebben. We passen daar noodgedwongen onze moderne criteria op toe. Het meer dan incidenteel voorkomen van een niet meer bestaande woordcombinatie als één woord wijst op lexicalisering in de desbetreffende taalfase; deze vorm moet dus in principe gehonoreerd worden. De samenstellende elementen worden bepaald bij de morfologische analyse, waardoor de gebruiker de relatie kan vinden tussen verbinding en samenkoppeling.

4.1 Samenkoppelingen

Een samenkoppeling die eertijds één woord was, kan nu nog één woord zijn:

mnl., nnl. niettemin	NIETTEMIN
mnl. mijnvrouwe	MEVROUW

maar ook niet meer als koppeling bekend zijn. We lemmatiseren dan toch de koppeling, omdat het geen incidentele clisisvorm is, maar als woord functioneerde.

vmnl. tsarmeer (te jare meer ‘voortaan’)	TSAARMEER
--	-----------

Ook als dit woord nog als verbinding bekend is, wordt eventuele assimilatie niet opgelost:

mnl. wattan (bijwoord en voegwoord)	WATTAN (niet: WAT + DAN)
mnl. harentare	HARENTARE (niet: HER + EN + DER)

De historische koppelingen vertonen vaak assimilatie en andere klankwijzigingen. Dit is zeker het geval bij bastaardvloeken; ook deze worden als de gestandaardiseerde oude vorm gelemmatiseerd en niet geanalyseerd.

mnl. gangswonden	GANSWONDEN (niet: GANS/GOD + WONDEN)
mnl. wetekeye	WETEKEE (niet: WETEN + CHRISTUS)
nnl. tjacob	TJACOB (onduidelijke status van de <i>-t</i>)

De vorm van het lemma bij een samenkoppeling de moderne vorm, als die bestaat:

asjebliëft	< als het je belieft	ALSJEBLIEFT
mod.nl. a.u.b.		ALSTUBLIEFT

Voegwoorden en bijwoorden die uit twee elementen bestaan, soms los, soms aaneengeschreven, worden als gewone samen koppelingen opgevat (anders dus dan uitgebreidere verbindingen; zie 4.4):

mnl., vnnl. als dat, alsdat	ALSDAT
mnl., vnnl. so dat, sodat	ZODAT
mnl., vnnl. te waren, te waar, tewaren	TEWAREN
mnl., vnnl. te hant, tehant	TEHAND
mnl., nnl. te hants, tehans, thans	THANS

Samen koppelingen die in lopende tekst los geschreven zijn, krijgen evenzo de moderne lemmavorm:

vnnl. aller liefst	ALLERLIEFST
vnnl. boven al	BOVENAL
nnl. wel bespraakt	WELBESPRAAKT

Hetzelfde geldt voor samen stellingen die ook niet-continu kunnen voorkomen, zoals scheidbaar samengestelde woorden: werkwoorden met een bijwoord als eerste element:

nnl. hij <u>belde</u> me <u>op</u>	OPBELLEN
------------------------------------	----------

en voornaamwoordelijke bijwoorden:

nnl. <u>daar</u> heb ik geen zin <u>in</u>	DAARIN
--	--------

Sommige lemma's bevatten een koppelteken, dat als verbindingsteken binnen één woord dienst, conform de moderne spellingregels: EX-VERSLAAFDE, GARAGE-INGANG, WC-BRIL, ZWART-WITTELEVISIE.

4.2 Meerwoordsexpressies

Inheemse vaste verbindingen, zoals idiomatische uitdrukkingen, worden in traditionele lexicografische bronnen bij een van de hoofdwoorden beschreven. Ook in GiGaNT krijgen de afzonderlijke woorden elk een eigen lemma. Het kan echter voorkomen dat dergelijke meerwoordsexpressies in het lexicon ook als ingang opgenomen worden, namelijk wanneer daar in een gekoppelde lexicale bron aanleiding toe is. Een verbinding die frequent als zinsdeel voorkomt en (meestal) aan elkaar geschreven wordt, is een samen koppeling, waaraan uiteraard één lemma wordt gehecht (zie hierboven, 4.1). Dat betekent dat afzonderlijke lemma's dus gelden voor:

mnl. tote dier stont dat	TOT / DE / STOND / DAT
mnl. hare no dare	HER / NO / DER

Vgl. de equivalente samen koppeling *harentare* onder 4.1.

Vreemde woorden die een verbinding vormen die als zinsdeel in het Nederlands functioneert vormen daarentegen altijd één lemma; de elementen hebben namelijk geen zelfstandige status in de Nederlandse woordenschat. Zie, volgens de moderne spellingregels gespeld: AGNUS DEI, VIS À VIS, EX AEQUO en zo ook:

nnl. Gloria Patri	GLORIA PATRI
-------------------	--------------

Gloria Patri is de naam van een gebed, die bestaat uit de beginwoorden. Enkele andere voorbeelden zijn:

pro Deo	PRO DEO
nnl. p.m., pro memorie	PRO MEMORIE
nnl. t. à t., tout à toi	TOUT À TOI

Soms worden dergelijke verbindingen als samen koppeling opgevat, zoals in het volgende geval. Het is de naam van de eerste zondag na Pasen.

mnl., nnl. quasi modo, quasimodo	QUASIMODO
----------------------------------	-----------

Een tweede categorie van meerwoordsexpressies die altijd één lemma vormen zijn eigennamen (zie ook hoofdstuk 8) die uit meerdere woorden bestaan. Vaak gaat het om geografische namen en meerdelige voor- of achternamen. De elementen van de naam worden verbonden, en ze krijgen samen één lemma.

Den Hage	DEN HAAG
Van Kerkhof	VAN KERKHOF
Piet Hein	PIET HEIN

4.3 Clitische combinaties

Bij incidentele clitische combinaties worden de moderne lemmagrenzen aangehouden.

mnl. nien - dat ic <u>niene</u> quam	NIET + NE
mnl. hine - dat <u>hine</u> (= hij hem) sloech	HIJ + HIJ

Bij clitische combinaties en assimilatie in verbindingen wordt het resultaat als samen koppeling opgevat als de vorm regelmatig voorkomt, en krijgt dan een overeenkomstig lemma met inachtneming van de spellingregels. Zie hiervoor verder onder 4.1.

mnl. mettesen	METTEZEN
mnl. tenengader	TENENGADER

4.4 Afkortingen

Hieronder zijn inbegrepen symbolen, letterwoorden, initiaalwoorden, verkortingen.

Afkortingen in historische lemma's worden opgelost. Wanneer de afkorting onduidelijk is, krijgt deze meerdere lemma's.

ct., cts.	CENT
l., LB, lib	LIBRA
mnl. d, den	DENARIUS DENIER

Toelichting: DENARIUS, wegens de spellingen *d*, *den*, *denarj*. voor lat. *denarius*; DENIER is de Franse vorm.

Afgekorte woorden krijgen het lemma van het volledige woord.

vnnl. aangen.	AANGENAAM
mod.nl. wed.	WEDUWE
mod.nl. j.l.	JONGSTLEDEN

Afkortingen die op verbindingen teruggaan, worden als clitische combinaties opgevat en als zodanig behandeld. Zie ook onder 4.3.

vnnl. ul (uwe liefde)	UW + LIEFDE
mod.nl. m.a.w.	MET + ANDER + WOORD
mod.nl. H.M.	HAAR + MAJESTEIT

Er zijn afkortingen en letterwoorden die zodanig gestandaardiseerd zijn dat ze als lemma de afkorting zelf krijgen. Wanneer de corresponderende voluit geschreven vorm gangbaar is en in GiGaNT is opgenomen, worden deze aan elkaar gekoppeld via eenzelfde superlemma-ID.

mod.nl. w.c., wc		WC
mod.nl. w.w., ww (werkloosheidswet)		WW
mod.nl. havo		HAVO

Verkorting zijn dikwijls ingeburgerd en krijgen dan die vorm als lemma:

mod.nl. auto	< automobiel	AUTO
mod.nl. bieb	< bibliotheek	BIEB
mod.nl. tuurlijk	< natuurlijk	TUURLIJK
mnl., nnl. spijt	< despijt	SPIJT

4.5 Foutieve vormen

Vormen die per vergissing zijn ontstaan, zoals kopieerfouten, drukfouten, typfouten en spelfouten, komen zowel in handschriften als in drukwerken voor.

Hiervoor geldt het onderscheid:

- evident fout, gevallen waarvan de correcte lezing duidelijk is, en die als lemma de correcte vorm krijgen. Deze worden gemarkeerd als foutief.

mod.nl. botcr (lees: boter)	BOTER
-----------------------------	-------

- niet te interpreteren; mogelijk corrupt/fout. Hier is het lemma conform de aangetroffen vorm.

mnl. <i>savret</i>	SAVRET
--------------------	--------

- vermoedelijk corrupt/fout, waarbij een alternatief kan worden gepresenteerd. Deze vorm kan een dubbel lemma krijgen; namelijk lemma 1 = lemma voor de corrupte vorm; lemma 2 = omspelling van het alternatief (het lemma is ambigu).

mnl. bedeesd	BEDEESD BETEESD
--------------	-------------------

(“waarschijnlijk bedorven lezing voor *beteest*.”)

5 De weergave van affixen

Affixen worden altijd in hun eigen vorm in het lemma weergegeven (bijvoorbeeld in een verkleinwoord, een afgeleide vrouwelijke persoonsnaam, een frequentatief van een werkwoord). Verschillen in woordstructuur zien we dus altijd terug in de lemmavorm. (Zie ook de voorbeelden in 2.5).

5.1 Handhaving van oorspronkelijke affixen

Verdwenen affixen worden ook opgenomen en niet vervangen door moderne:

mnl. weghelijn	WEGELIJN (niet: WEGJE)
----------------	------------------------

(zie voor diminutiefsuffixen verder onder 9.5.2.5).

mnl. erfacht	ERFACHT (niet: ERFACHTIG)
mnl. broekacht	BROEKACHT (niet: BROEKIG)

Over dit *-acht* is het WNT niet zo duidelijk, maar zie EWN [-achtig](#) en Pijnenburg (1993).

-el wordt als affix beschouwd; het komt onder andere voor in gereedschapsnamen in:

mnl. seisel, sessel	ZEISEL
---------------------	--------

Daarentegen:

mnl. seisen, seisene:	ZEIS
-----------------------	------

want de status van *-en* is hier onduidelijk.

5.2 Standaardvormen

Voor affixen zijn standaardvormen vastgesteld. Zie voor specifieke gevallen 9.5.2.2 t/m 9.5.2.5.

onl. diupitha	DIEPTE (niet: DIEPEDE)
---------------	------------------------

mnl. drapenierigghe	DRAPENIEREGGE
---------------------	---------------

mnl., westvl. sonderlanghe	ZONDERLING
----------------------------	------------

Verschillende (klank-/spelling)vormen van structureel eenzelfde afleiding worden tot één vorm teruggebracht, namelijk de vorm die we in moderne afleidingen zien:

lengede, langde	LENGTE
-----------------	--------

ternink, teerling	TEERLING
-------------------	----------

Voor affixen die een verschillende standaardvorm hebben, wordt per woord beslist welke standaardvorm het affix in het lemma krijgt. In het lexicon worden eventueel nog gangbare varianten ook opgenomen.

Voorbeeld 1: *-erie* en *-erij*.

rabauwerie	RABAUWERIJ
------------	------------

clownerie	CLOWNERIE
-----------	-----------

selderij	SELDERIJ, SELDERIE, SELLERIE
----------	------------------------------

Voorbeeld 2: *-erd* en *-aard*.

mnl. mostaert, mostert	MOSTERD
------------------------	---------

mnl. ruwart, rewert	RUWAARD
---------------------	---------

mnl. tabbart, tabbert	TABBAARD, TABBERD
-----------------------	-------------------

Voorbeeld 3: *-aar* en *-er* worden niet als verschillende affixen beschouwd, en een oudere vorm op *-er* kan dus een suffix *-aar* krijgen:

mnl. <i>wanderer</i>	WANDERAAR
----------------------	-----------

Dit woord volgt de regels van het moderne Nederlands *-aar* na werkwoordstammen die eindigen op *sjwa + l, n, r*.

6 Lemmatisering van voornaamwoorden en telwoorden

Bij de lemmavorm voor voornaamwoorden laten we grammaticale eigenschappen prevaleren boven de fonetische en etymologische. Dit houdt in dat we één hoofdvorm voor een paradigma aanhouden. Voor de gebruiker zijn alle andere, ook qua wortel afwijkende, vormen van het paradigma vindbaar via de vormen die bij de morfosemantische kenmerken zijn opgenomen. Hieronder volgen beslissingen en bijzonderheden over de voornaamste categorieën. Ook wordt er kort aandacht aan lemmatisering van telwoorden besteed.

6.1 Persoonlijk voornaamwoord

Alle vormen in eenzelfde paradigma krijgen als lemma de subjectsvorm:

mnl., vnl. du, dijns, di DU

mnl., nnl. wi, wij, we, onzes, ons WIJ

Gereduceerde en clitische vormen krijgen de volle lemmavorm:

'r - hij heeft r (< haar) gezien ZIJ

-et - (hi) naemt (< hi nam-et) NEMEN + HET

Problematisch zijn vooral de voornaamwoorden van de tweede persoon, omdat er verschuivingen binnen de paradigma's hebben plaatsgevonden.

a. *u - gij*

U, oorspronkelijk alleen objectvorm van de tweede persoon en dus een vorm van het paradigma van *gij*, verkreeg hiernaast subjectfunctie, met overgangsvormen als *ue*, *uwee*. In principe zou aan objectvormen in het Nieuwnederlands t/m de 19^{de} eeuw een dubbel (ambigu) lemma toegekend moeten worden. Om pragmatische redenen wordt echter naar de vorm gelemmatiseerd.

mnl. u, ju, jou (altijd object) GIJ

nnl. u, ou, oe (object en subject) U

Een bijzonder geval is:

nnl. U.E., UE., UE, u E., ue. U ED.

Deze afkorting kan niet opgelost worden, omdat niet duidelijk is of zij voor *Uwe Edelheid* of *U Edele* staat. Ze kan in deze (korte) vorm als persoonlijk voornaamwoord opgevat worden.

Ook de klankvorm *uwee* krijgt deze lemmavorm. Zie verder WNT [UE](#).

nnl. uwee, uwé U ED.

nnl. uwees (nominativering bez.vnw. uwees) U ED.

b. *gij - jij*

Jij, oorspronkelijk nevenvorm van *gij*, kreeg in het Nieuwnederlands gaandeweg een eigen status als voornaamwoord voor informeel gebruik of voor aanspreking van minderen. Ook de objectvormen werden *j*-vormen. Zo verwijzen de lemma's *ju* en *jou* in het MNW naar de hoofdvorm *gij*, in het WNT wordt bij dezelfde vormen naar *jij* verwezen. Ook hier zou het mogelijk zijn t/m de 19^{de} eeuw een dubbel (ambigu) lemma toe te kennen. Om pragmatische redenen wordt ook hier naar de vorm gelemmatiseerd:

mnl. ge, gy, jou GIJ

nnl. gij, ge GIJ

nnl. jij, je, jou JIJ

c. *gijlieden - jullie*

De meervoudige pers. vnw. kennen uitgebreide vormen met *-lieden*, waarvan die van de tweede persoon, *gijlieden*, *jijlieden* bijvormen had als *jullie* en *jelui*. *Jullie* werd een nieuwe standaardvorm voor de 2^{de} persoon meervoud. Het WNT behandelt dit *jullie* onder het lemma *jij*, vanuit 19^{de}-eeuws perspectief, maar dit past niet bij het moderne voornaamwoordelijke systeem. Daarom worden alle vormen waarin het element *-lieden*, *-lieden* herkenbaar zijn als *-lieden*-vorm gelemmatiseerd, met, al naargelang de beginconsonant, *gij*- of *jij*- als eerste element. Overige vormen vallen onder JULLIE.

Lemmavoorbeelden:

gijlieden	GIJLIEDEN
jelieden, jeluider, julder	JIJLIEDEN
jijlui, joului, jelui, jullie	JULLIE

6.2 *Bezittelijk voornaamwoord*

Zij krijgen een afzonderlijk lemma, hoewel ze voortkomen uit de persoonlijke voornaamwoorden. Gereduceerde vormen als *m'n* en *je* worden gestandaardiseerd tot MIJN en JOUW. Er zijn zelfstandige gebruikte bezittelijke voornaamwoorden als: het *mijne*, de *zijnen*. In dat geval wordt de woordsoort zelfstandig naamwoord en wordt het lemma MIJNE respectievelijk ZIJNE.

6.3 *Aanwijzend voornaamwoord*

In historisch materiaal vallen deze voor een deel qua vorm samen met het lidwoord. De context is bepalend. Let wel, HET gaat DAT als lidwoord vervangen, daarbij is geen sprake van samenvval (zie 3.4).

6.4 *Reflexief voornaamwoord*

In ouder Nederlands (en nu soms nog in regionaal en informeel taalgebruik) vervult de objectsvorm van de voornaamwoorden deze functie. Daarom hebben woorden als *hem* en *haar* in deze functie wel een eigen lemmavorm. Ze krijgen dus niet de subjectvorm als lemma, en zeker niet ZICH.

haer - si rechte haer op	HAAR
'em - dat hij 'em vergist had	HEM

6.5 *Reciproque voornaamwoord*

Het moderne *elkaar* heeft als voorgangers *elk-ander* en *manlijk-ander*, die ook niet-continu kunnen voorkomen en vele verkorte en geassimileerde vormen hebben. Alle *elk*-vormen woorden als ELKANDER en alle *manlijk*-vormen als MALKANDER gelemmatiseerd. In het lexicon zijn ELKAAR-ELKANDER en MALKAAR-MEKAAR-MEKANDER te vinden als varianten.

6.6 *Onbepaald voornaamwoord*

Vooraf oudere vormen van onbepaalde voornaamwoorden kennen door verbuiging en combinatie met diverse woorden allerlei variantie. Doorgaans worden verbogen vormen niet afzonderlijk gelemmatiseerd, maar wordt de woordstructuur wel gehandhaafd. Als een woord nog bekend is, kan de huidige vorm wel invloed op het moderne lemma hebben.

Lemmavoorbeelden:

als 'alles', alles	ALLES
soghedaen	ZOGEDAAN (niet: ZODANIG)
diesgelijk(s), desgelijk(s)	DESGELIJK(S)
el 'ander'	EL

elkendeen	ELKENDEEN
letter, lettelt, luttelt	LUTTEL
negeen, egeen, engeen	NEGEEN
niemen, niemant, niemanden	NIEMAND
selck, zulk	ZULK
zulks	ZULKS

6.7 Telwoorden

Hoofdtelwoorden worden voluit geschreven bij getallen van nul tot en met honderd en bij honderdtallen en duizendtallen. Alle overige hoofdtelwoorden krijgen een getal zonder . of , als scheidingsteken voor getallen vanaf 1000. Jaartallen worden ook als getal gelemmatiseerd. Op deze manier worden verschillende weergaves van hetzelfde telwoord bij elkaar gehouden (bijvoorbeeld 1496 voor *veertienhonderdzesennegentig* en *duizend vierhonderdzesennegentig*).

honderd	HONDERD
honderdeen, honderdeneen, 101	101
vijfduizend	VIJFDUIZEND
mnl. vierwarf twintich (het jaar) 1962	VIERWERFTWINTIG 1962

Rangtelwoorden krijgen hun eigen lemma, waarbij het onderscheid tussen de suffixen *-de* en *-ste* wordt behouden. Wanneer in een corpus een rangtelwoord met een getal wordt aangegeven, en de data is 17e-eeuws of ouder, dan krijgt dat getal een dubbel lemma.

tweede	TWEEDE
vierste	VIERSTE
mnl. <i>Up den 2 dach van Septembre</i>	TWEESTE TWEEDE

Breuken bestaan uit een hoofdtelwoord gevolgd door een (eventueel gesubstantiveerd) rangtelwoord. Elk deel krijgt zijn eigen lemma. In aan het lexicon gekoppelde lexicale bronnen kan een breuk als (uitgeschreven) meerwoordsexpressie worden opgenomen.

$\frac{1}{2}$	EEN + TWEEDE
$\frac{1}{4}$	EEN + VIERDE
$\frac{1}{6}$	EEN + ZESDE

7 Leenwoorden

Leenwoorden met uitheemse klanken en spellingen krijgen, als ze nog bekend zijn, de moderne vorm als lemma. Soms gelden daarnaast specifieke regels.

7.1 Aansluiting bij de oorspronkelijke vorm

Woorden afkomstig uit een vreemde taal in hun oorspronkelijke vorm, voor zover tenminste als Nederlands woord gebruikt, krijgen de standaardvorm uit die taal.

Voorbeeld 1:

npl. intrade ‘intoct’ (Spaans, Italiaans)	INTRADE
npl. negligé (Frans)	NEGLIGÉ

Als een woord in het Nederlands merendeels aansluiting laat zien bij de oorspronkelijke vreemdtalige vorm, dan wordt deze zo overgenomen.

Voorbeeld 2, bij fr. *aise*:

mnl. eise	AISE
mnl., nnl. asement, aisement	AISEMENT
mnl. aaisieren	AISEREN

7.2 Meervoudige ontleningen

Leenwoorden die in een verschillende periode en/of via verschillende talen ontleend worden, en naast vormverschil ook verschillende betekenissen kunnen hebben, worden als verschillende woorden opgevat en krijgen verschillende lemma's.

Voorbeeld 1, beide uit Frans *occasion*:

mnl. occasioen 'gelegenheid'	OCCASIOEN
mod.nl. occasion 'koopje; tweedehandsauto'	OCCASION

Voorbeeld 2:

mnl., nnl. trop < lat. troppa	TROP
nnl. troep < fr. troupe	TROEP

Voorbeeld 3:

mnl., nnl. foelge, foelie 'dun blad'	FOELIE
nnl. folie 'blad uit boek; dun vel'	FOLIE
nnl. folio 'blad uit boek; formaat'	FOLIO

Toelichting: folio is ontleend uit het Latijn, foelie via het Frans uit het Latijn, folie via het Duits uit het Latijn. Daarnaast is FOLIE 'dwaasheid' een homoniem naast FOLIE 'dun vel'.

Voorbeeld 4:

mnl. abattement < fr. <i>abattement</i>	ABATTEMENT
nnl. esbattement < fr. <i>e(s)battement</i>	ESBATTEMENT

7.3 Verouderde vreemde spellingen

Deze worden niet gehonoreerd. Leenwoorden als mnl., vnnl. *bataelge*, *bottelgier*, *selgeren*: de -g- hierin is een weergave van de gemouleerde -l- in het Frans; die spelling verdwijnt in de loop van de geschiedenis (Salverda de Grave (1906: 279). De lemmavormen hebben daarom altijd -l- of -ll-.

mnl. selgeren (asselgeren, asselieren)	SELEREN
mnl. faelgieren	FAILLEREN

Dit geldt ook voor affixen: het Franse suffix *-age* dat in het oudere Nederlands ook als onder andere *-aedse* en *-aige*, *-adze* voorkomt, wordt in het lemma steeds als *-age* weergegeven, zoals in:

mnl. dijckaedze, dijckaetge, dijckaidzen	DIJKAGE
mnl. stallaedse, stalaigen, stallagen	STALLAGE

8 Eigennamen

Namen van personen (voornamen en toenames/familienamen) en toponiemen (dorpen, steden, rivieren, percelen, landen en dergelijke) maken van elke tekst deel uit. Naast deze hoofdcategorieën van eigennamen zijn er nog namen van andere entiteiten zoals organisaties, gebouwen en schepen. Voor de lemmavorm van eigennamen gelden regels die kunnen afwijken van die voor de soortnamen. Hieronder volgt een nadere toelichting.

8.1 Relatie tot soortnamen

De historische namenvoorraad is bij het toekennen van standaardvormen te onderscheiden in diverse categorieën voor wat hun aansluiting bij het niet-onomastische Nederlands betreft:

1. inheemse namen die sterk tegen de appellatieven aanleunen: van de persoonsnamen zijn dat tot aan de moderne tijd enkele typen toenames (zoals beroeps(bij)namen als *de bakker* en eigenschapsnamen als *witkop*), van de toponiemen bijvoorbeeld de inheemse plaatsnamen en perceelnamen. Voorts kunnen dat namen van zaken als organisaties en gebouwen zijn. Overigens kunnen deze namen zelf weer zuiver onomastische elementen bevatten: *Jans stuk* (perceelnaam met daarin voornaam), *van Antwerpen* (toenaam met daarin toponiem);
2. inheemse namen die meer een naamkarakter dragen, doordat het niet zonder kennis van de etymologie duidelijk is dat ze componenten bevatten die van oorsprong wel een appellatief waren: voornamen (*Geertrui*); familienamen in de moderne tijd (*De Vooys*);
3. vreemde namen: uitheemse antroponiemen, gelatiniseerde inheemse namen en toponiemen zonder relatie met Nederlandse soortnamen (*Adam, Indië*).

8.2 De werkwijze per categorie

8.2.1 Inheemse (Germaanse) voornamen

Deze bevatten vaak twee elementen die in het eerste lemma min of meer gestandaardiseerd worden. Voorbeelden zijn *-bert/-brecht*, *her-*, *ijs-*, *-rik*, *-ger/-geer*, *-mar*, *-burg*, *-zoete*. Er zijn ook eenstammige namen en verkorte vormen.

Opmerkingen:

- a. Vormen die slechts weinig variatie vertonen worden samengenomen.
- b. Gelatiniseerde en geromaniseerde vormen van Germaanse namen krijgen een eigen lemma. Daarom worden vormen op *-e*, *-a* en *-o* onderscheiden.
- c. Vleivormen en verkleinvormen krijgen hun eigen lemma; de vleisuffixen en diminutiefsuffixen hebben een standaardvorm: *-zoete*, *-man*, en *-tjel-je* (uit *-kin* etc.), *-ijn*, *-sijn*, *-lijn*.

gherbrecht	GERBRECHT
aleburgh	AALBURG
heinin	HEINIJN
heinekin	HEINTJE
henc	HENK

henrick	HENRIK
heinricus	HEINRICUS [gelatiniseerde vorm]
aleit	ALEID
adelisa	ADELISA
aleidis	ALEIDIS [gelatiniseerde vorm]
adelheid	ADELHEID
addin (1e lid mogelijk <i>adel-</i> , 2e lid <i>-ijn</i>)	ADDIJN
boetsoete	BOETZOETE

8.2.2 Uitheemse voornamen en andere persoonsnamen

Ingeburgerde uitheemse namen hebben veelal een aan het Nederlands aangepaste vorm, waarop het lemma wordt gebaseerd. Hier kunnen verkortingen en vleivormen bij zijn. Zo mogelijk berust de volle vorm op de naamvorm van de geïdentificeerde persoon.

griet	GRIET
piterken	PIETERTJE
lisebette	LIESBET
janne	JAN
germain	GERMAIN
femine	FEMINE
iram	IRAM
facee	FACEE
fierin	VIERIJN
kerst	KERST (niet: CHRIST of CHRISTUS)
apollo	APOLLO
apollin	APOLLIJN

Sommige namen worden uit verschillende talen ontleend. Dit onderscheid komt terug in het lemma.

Hektor	HEKTOR [Griekse vorm]
Hector	HECTOR [Latijnse vorm]

Voor de lemmatisering van Bijbelse namen is de recente lijst van het Nederlands Bijbelgenootschap (<https://www.debijbel.nl/bijbelse-namen>) leidend. Als echter de vorm uit deze lijst sterk afwijkt van de historische vorm, wordt als lemma een vorm gekozen die bij de laatste aansluit.

8.2.3 Appellatieven als toenaam

Zeker in het Oud- en Middelnederlands is bij toenames het onderscheid tussen overerfbare namen en individuele namen niet te maken. In beide gevallen is nog geen sprake van een vaste vorm en daarom worden deze als appellatieven gelemmatiseerd.

vmnl., mnl. hoenredief	HOENDERDIEF
onl. pankuoko	PANNENKOEK
vmnl. de bakker	DE / BAKKER

8.2.4 Moderne namen

Hoewel de appellatieven in moderne namen vaak herkenbaar zijn, worden de namen vanaf de moderne tijd niet meer als appellatieven gelemmatiseerd, en de spelling ongewijzigd uit de bron overgenomen. In geval dat uit de bron duidelijk is dat ze naar eenzelfde referent verwijzen, worden ze wel geüniformeerd. Voornamen en familienamen die uit meer woorden bestaan worden als één vorm gelemmatiseerd, bij andere namen krijgt elk woord een eigen lemma.

de Ruijter	DE RUIJTER
Van der Heijden	VAN DER HEIJDEN
(Peter) Pannekoek	PANNEKOEK

8.2.5 Inheemse toponiemen

Historische toponiemen bevatten vaak overgeleverde vormen die als appellatief gezien kunnen worden (en die herkenbaar zijn in de standaardlemma's van de historische woordenboeken). Daarnaast hebben de meeste een gefixeerde, moderne naam die de plaats identificeert. Soms heeft de moderne naam gedeeltelijk (zoals *Oegstgeest*) of geen (zoals *Monsterhoek*) taalkundige relatie met de oude naamvorm.

Voorbeeld 1: de plaatsnaam *Oegstgeest*:

osgeresgest (opgetekende vorm 10^{de} eeuw) 'geest(grond) van Osgeer'; *oesghest*, *oesgeist*, *oesgheest* (opgetekende vormen 13^{de} eeuw): *oegstgheest*: latere vormen met herinterpretatie van *oest* > *oegst*, *oogst*.

Oegstgeest: de huidige naam. Omdat het element *Oegst* zich heeft ontwikkeld uit *Osgeer* wordt het lemma ook altijd OEGSTGEEST.

Voorbeeld 2: de poldernaam *Monsterhoek* (polder bij Krabbendijke op Zuid-Beveland, Zeeland):

monsterhoec (opgetekende vorm 13^{de} eeuw) 'kloosterhoek (stuk land van/bij klooster)'.
Monnikenpolder: de tegenwoordige naam.

Het lemma van *monsterhoec* wordt MONSTERHOEK, want het is een andere benaming voor dezelfde plaats.

Toponymische elementen worden in de lemma's gestandaardiseerd: *-dijk*, *-lo*, *-voorde*. Bij bepaalde elementen wordt ook de gereduceerde vorm of vormvariant als lemmadeel opgenomen (*-em*, *-heem*, *-heim*).

dorremael	DORMAAL
achtienhoven	ACHTTIENHOVEN
scaepettinc	SCHAAPETTING
boekhout (twee verschillende plaatsen)	BOEKHOUT óf BOEKHOUTE

Een variante naam voor dezelfde plaats krijgt een eigen lemma.

Den Haag	DEN HAAG (niet: 'S-GRAVENHAGE)
----------	--------------------------------

Plaatsnamen kunnen van de normale spellingsregels afwijken. Vergelijk NIEUW-VENNEP (plaats) en OUD-BEIJERLAND (plaats) conform de spellingregels naast NIEUW VENNEP (naam station) en OUD GASTEL (plaats).

8.2.6 Uitheemse toponiemen en toponiemen met uitheemse elementen

Uitheemse namen en namen die uitheemse elementen bevatten, behouden hun eigen lemma. Is er een Nederlandse benaming hiervoor, dan krijgt die een eigen lemma (zie bijvoorbeeld de [Buitenlandse aardrijkskundige namen in het Nederlands](#)).

Plymouth	PLYMOUTH (Engelse vorm)
Pleimuiden	PLEIMUIDEN
Torino	TORINO (Italiaanse vorm)
Turijn	TURIJN
Slovensko	SLOVENSKO (Sloveense vorm)
Slowakije	SLOWAKIJE, <i>SLOVAKIJE</i>

Is de oude naam vervangen door een nieuwe naam, dan krijgt het toponiem een lemma conform de oude vorm.

Ceylon	CEYLON (niet: SRI LANKA)
Siam	SIAM (niet: THAILAND)

9 Detailproblemen en voorbeelden bij 3 t/m 5

Bij 3 Het lemma en eventuele varianten

9.3.1 Het al of niet volgen van regelmatige klankveranderingen

Bij verdwenen woorden geldt de regel voor klankverandering die algemeen heeft gewerkt (zie hierboven), bijvoorbeeld *e*-apocope:

are ‘arend’	AAR
-------------	-----

of diftongering en *e*-apocope:

mnl. mute ‘ruien van vogels, ruitijd’	MUIT
---------------------------------------	------

Is een woord nu niet meer bekend, maar wel opgenomen in het WNT in de spelling De Vries en Te Winkel (op basis van de “beschaafde uitspraak” van de 19^{de} eeuw, dat wil zeggen de klankvorm van de toen beoogde standaardtaal), dan geldt die spelling/vorm.

Voorbeeld 1, wisseling o - eu:

mnl. uteboren	UITBEUREN
nl. uitbeuren	UITBEUREN

Bepaalde klankwijzigingen hebben maar in een deel van de woordenschat gewerkt. In etymologisch nauw verwante woorden kan ook het ene de klankwijziging vertonen en het andere niet: *bruisen* (met *oe* > *ui* en diftongering) *tgov. broes* ‘tuit van een gieter’ (*oe*-relict); *vouwen* (met *d*-syncope) *tgov. eenvoudig* (zonder *d*-syncope). Voor deze woorden hanteren we de moderne variant, als die nog bestaat. Zo niet, dan construeren we de meest regelmatige vorm:

Voorbeeld 2:

vuldelic, voldelijk	VOUDELIIJK
---------------------	------------

namelijk met regelmatige overgang *old* > *oud*, zoals nog in *eenvoudig*, en niet met *d*-syncope, zoals in *vouwen*.

Soms is een recente vorm zo zeldzaam (bijvoorbeeld alleen dialectisch) dat de gewonere Middelnederlandse variant de standaard wordt.

Voorbeeld 3:

onl., mnl. horik, hornik	HORNIK
nnl. hurk (sporadisch aangetroffen)	HORNIK

Voorbeeld 4: Als er frequente samenstellingen zijn met een bepaald woord dat een variant heeft, zoals woorden op *-ceel/-cedel* ‘bewijsstuk’ en het blijkt dat de vormen ook binnen de samenstellingen zelf vaak variëren, kan het beter zijn één hoofdvorm in het lemma te kiezen, in dit geval de *d*-vorm *-CEDEL* (voor de regel dat de *-d*-vorm vóórgaat (zie 9.3.3., syncope van *-d-*). Ook hier is een uitzondering: het lemma is *DOOPCEEL*, omdat deze variant nog bekend is in de uitdrukking *iemands doopceel lichten*.

Soms speelt desambiguering een rol. Vandaar uniformering op *-gaard* ‘tuin’ bij vele samenstellingen, echter met afwijking van *DIERGAARDE* bij *GAARD*, conform de moderne vorm (zie ook bij 3.3 Desambiguering).

Varianten van een bepaalde afleiding of samenstelling worden hierbij natuurlijk wel steeds door hetzelfde lemma aangeduid. Dus geen lemma *DOOPCEDEL* naast een lemma *DOOPCEEL*.

9.3.2 Enkele specifieke klankwijzigingen 1: apocope van *-e*

In het moderne Nederlands is de eind-*e* (sjwa) veelal verdwenen. Die wordt dan ook niet in het lemma opgenomen. Zie bekende woorden als *ster*, *vrouw*, *veen*, *kerk* maar ook onbekende woorden als

are ‘arend’	AAR
-------------	-----

De apocope-regel geldt ook voor vreemde woorden waarvan een *e*-loze vorm is aangetroffen:

Voorbeeld 1:

mnl. <i>daet</i> (< fr. <i>date</i> , lat. <i>datum</i>)	DAAT
mnl., nnl. <i>date</i>	DAAT

Zo ook *fore* met lemma *FOOR*.

Soms zijn er redenen om de *-e* te behouden:

Voorbeeld 2:

mnl., nnl. <i>miede</i> ‘loon’	MIEDE
--------------------------------	-------

hoewel de recentste citaten van dit verouderde woord ook wel eens *mie* of *mied* hebben. Immers, de *d*-vorm moet altijd toegekend worden; bovendien zijn *MIE* en *MIED* al andere lemma’s, vandaar behoud van *-e* (desambiguering).

Persoonsaanduidingen op *-e* (m/v) die gesubstantiveerde adjectieven zijn, behouden hun uitgang:

mnl., nnl. <i>blinde</i> , <i>blinden</i>	BLINDE
mnl. <i>fine</i>	FIJNE
nnl. <i>de Allerhoogste</i>	ALLERHOOGSTE

Hetzelfde geldt voor zelfstandig gebruikte adjectieven die een onzijdig zelfstandig naamwoord opleveren, zij het dat deze in aan GiGaNT gekoppelde lexica bronnen niet altijd opgenomen zijn:

nnl. dat alrebeste

ALLERBESTE

nnl. het grappige van die zaak

GRAPPIGE

NB: Eind-*e*'s die een suffix zijn, blijven bewaard; zie hiervoor onder hoofdstuk 9.5.2.2, Affixen.

9.3.3 Enkele specifieke klankwijzigingen 2: syncope van -d-

Dit verschijnsel levert veel varianten op, omdat vaak zowel de vormen met als die zonder -*d*-syncope in het Standaardnederlands zijn terechtgekomen (*leder én leer*), anders dan de varianten met de overgangsklanken -*j*- en *w*-, waarvan de oorspronkelijke vorm met -*d*- bijna altijd de meest geaccepteerde vorm is: *raden* vs. *raaien*; *houden* vs. *houwen*.

Een eerste regel is: bij het simplex van woorden waar *d*-syncope optreedt wordt voor het lemma altijd de vorm met -*d*- gekozen voor de corpustagging, waarbij in het lexicon als lemmavariant de vorm met syncope wordt opgenomen als deze in het hedendaags Nederlands voorkomt. Bij de lemmatisering moet verder rekening gehouden worden met homoniemen: LEER 'ladder' en LEER 'doctrine' hebben geen *d*-vormen zoals LEER/LEDER 'dierenhuid', evenmin als WEER 'man' en WEER 'inrichting voor visvangst', tegenover WEER/WEDER 'atmosferische toestand' en WEER/WEDER 'terug, nogmaals'. Ook dit pleit ervoor *d*-vormen toe te kennen waar die etymologisch thuishoren (desambiguering).

Voorbeeld 1, blode-blo (met een vorm *bleu*, die ook in de standaardtaal terecht is gekomen):

nnl. blode

BLODE, *BLO*, *BLEU*

nnl. bloode, bloo, blo

BLODE, *BLO*, *BLEU*

Voorbeeld 2:

nnl. leder, leer 'dierenhuid'

LEDER, *LEER*

In de samenstellingen kan er verschil zijn in behoud van de -*d*- in het moderne lemma. Zo zijn er nieuwe woorden die (waarschijnlijk) nooit een "oude" vorm gehad zouden kunnen hebben. Andersom zijn sommige oudere vormen met -*d*- in die vorm gefixeerd:

nnl. weder 'opnieuw' 'luchtgesteldheid'

WEDER, *WEER*

nnl. weerstation

WEERSTATION

nnl. wedergeboorte

WEDERGEBOORTE

Voorbeeld 3: dezelfde regel geldt ook voor hypercorrecte *d*-vormen die nu algemeen Nederlands zijn geworden: *kade* naast ouder *ka* en *kaai* < fr. *caye* (zie Schönfeld, paragraaf 37). Alles krijgt het lemma *KADE* en daarnaast *KAAI* en *KA*, althans in het simplex. Bij de samengestelde woorden komt in het WNT veel *kaai* voor, de oudere fasen hebben ook *ka* (*kadijk*).

Voorbeeld 4:

nnl. vader, vaar

VADER, *VAAR*

nnl. broeder, broer

BROEDER, *BROER*

Er zijn hier moderne vorm- en betekenisverschillen: *broeder* 'zoon van zelfde ouders' (formeel), 'verpleger', 'kloosterling' (neutraal), naast *broer* 'zoon van zelfde ouders' (neutraal). (Hetzelfde speelt bij *zus/zuster*.) Het is niet mogelijk die betekenisverschillen altijd te honoreren. Bij *broer/broeder* is de volgende lijn gevolgd: in principe is voor *BROEDER*, *BROER* (met extra lemma in het lexicon) gekozen in het simplex en in samengestelde woorden die ook

nu de *-broer*-variant kennen, zoals HALFBROER, STIEFBROER, SCHOONBROER. Enkele idiomatische samenstellingen krijgen alleen -BROER- : DRINKEBROER, JABROER. Verder steeds -BROEDER-. Bij *vader* behouden we de d-loze vorm als lemmavariant in het lexicon en in samengestelde woorden als BESTEVAAR. Verder steeds -VADER-.

Voorbeeld 5: *vlade/vla/vlaai*. Betekenissen en vormen lopen hier door elkaar. De oorspronkelijke woordvorm is *vlade* met de betekenis ‘dunne koek’. In het moderne Nederlands bestaat de vorm-betekenisverdeling: VLADE ‘dunne koek’, VLAAI ‘speciale (Limburgse) taart’ en VLA ‘koude gladde pap’. Hieronder de historische en moderne woordvormen met de belangrijkste betekenissen:

mnl. vlade, vlaey ‘dunne koek’	VLADE
nnl. vlade, vlaai, vla ‘koek; pap; gebak’	VLADE, VLAAI, VLA
mod.nl. vla ‘koude, gladde pap’	VLA
mod.nl. vlade ‘dunne koek’	VLADE
mod.nl. vlaai ‘gebak; koeienuitwerpsel’	VLAAI

In de samengestelde woorden komen we alle drie de varianten tegen.

9.3.4 Onregelmatige en niet goed traceerbare klank- en vormwijzigingen

Deze leiden niet altijd tot een afzonderlijk lemma.

Voorbeeld 1: de dialectvormen *wessel-wissel-wiezel(tje)-weizel(ke)*, door het WNT alle beschouwd als vormen van *wezel* ‘klein roofdier’, krijgen allemaal ook het lemma WEZEL.

Voorbeeld 2: gevallen van (incidenteel optredende) dissimilatie: de moderne vorm is standaard.

mnl. camerier	KAMENIER
mnl. ruggelde ‘gerugd’	RUGD (anorganische <i>-l-</i> uit <i>ruggede</i>)
mnl. edeken ‘herkauwen’	EDIKEN

Voorbeeld 3, vocaalvarianten:

Soms zijn *i-* en *e-*vormen varianten van elkaar maar niet altijd:

mnl. sencken ‘doen zinken’	ZENKEN (niet: ZINKEN)
----------------------------	-----------------------

Toelichting: dit werkwoord heeft ablaut met causatieve functie; vgl. *drenken*.

rinnen	RINNEN (niet: RENNEN)
--------	-----------------------

Toelichting: hoewel dit werkwoord verouderd is, en de betekenis deels door *rennen* overgenomen, krijgt het een eigen lemma wegens eigen paradigma (sterk tegenover zwak) en eigen betekenisontwikkeling.

Voorbeeld 4:

Als in het moderne Nederlands alleen de gewijzigde vorm nog te vinden is, wordt niet de etymologische vorm gelemmatiseerd.

mnl., nnl. achterskind	ACHTERSKIND (niet: ACHTERSTKIND)
gesternte	GESTERNTE (niet: GESTERTE)

9.3.5 Gewone tegenover gemarkeerde varianten in de moderne standaardvorm

De vorm die in het moderne Nederlands het gebruikelijkst is, wordt lemma.

Voorbeeld 1: *calander, kalander*. In de woordenboeken *kalander* (gewone vorm) en *klander* (gemarkeerde vorm). Het moderne lemma wordt KALANDER.

Voorbeeld 2: Bij minder bekende woorden is een frequente historische spelling soms belangrijker dan de vorm in moderne woordenboeken/-lijsten. Het woord *kateil* heeft historisch vooral *ei*-spellingen. Daarom is het lemma KATEIL geworden.

9.3.6 Volksetymologische wijzigingen en woordspelingen

Volksetymologische wijzigingen worden in het lemma gehonoreerd, we gaan dus niet “retrolemmatiseren”.

Voorbeeld 1:

Bij mnl. *schaerwetter*, met *-wetter* < *aweit* e.d., Frans leenwoord, is het lemma SCHAARWETTER (*wetter* ‘natmaker’).

Voorbeeld 2:

mnl., nnl. *aarzelmaand* ‘oktober’ kan afgeleid worden van het werkwoord *aarzelen*. Ook al noemen VMNW en MNW de mogelijkheid van vorming met *herse-* ‘gierst’, het lemma is die conform de aangetroffen vormen: AARZELMAAND.

Voorbeeld 3:

Volksetymologische vervormingen van leenwoorden krijgen hun eigen lemmavorm.

pontificaal	PONTIFICAAL
proncke-fykaal	PRONKFICAAL
sponceficael	SPONSFICAAL

Dat betekent dat naast *ponteficaal* er een lemma *pronkficaal* en *sponsficaal* bestaat. Het tweede element wordt gelemmatiseerd als -FICAAL, zoals in het oorspronkelijke woord *pontifical*.

Voorbeeld 4:

mnl. warant ‘borg, garantie’	WARANT
mnl. waerhant ‘borg’	WAARHAND

Voorbeeld 5:

Hierbij horen ook opzettelijke “creatieve” analogische varianten (woordspelingen): diverse *war*-varianten van ‘wereld’, die refereren aan het begrip ‘war’, ‘warrig’ krijgen het lemma WARRELD. En naast *legende* (lemma: LEGENDE) treffen we ook *leugende* (lemma: LEUGENDE) aan. NB: Dit type is te onderscheiden van verbasteringen die niet gemotiveerd zijn door betekenis of vormanalogie (al is het onderscheid soms moeilijk te maken):

Voorbeeld 6:

nl. variabel	VARIABEL
nl. wittelje, wittille, wittellie	WITILLIGE (niet gemotiveerd)

naast *wittellie* en *witheilige*, die niet het lemma WITILLIGE krijgen maar WITTELELIE respectievelijk WITHEILIGE.

9.3.7 Morfologie: huidige regels worden gevolgd

Voorbeeld 1:

Volgens het Morfologisch Handboek (De Haas en Trommelen 1990: 245) wordt voor het suffix *-(e)nis* de *-v-* waar de werkwoordstam op eindigt, stemloos (*droefenis*, *erfenis*). Bij *-(e)lijk* is dat soms ook het geval: *erfelijk*. Daar geldt het ook voor de *z*: *vreselijk* (De Haas en Trommelen 1990: 296).

mnl. lovenesse	LOFENIS
mnl. erloofnesse	ERLOFENIS
mnl. gelofnesse (ook met korte vocaal)	GELOFENIS

Voorbeeld 2:

d-epenthese na *-r-* bij de comparatieven van adjectieven en bij znw. op *-er*: *raarder*, *spaarder*. Deze *-d-* treedt in historische vormen ook na andere consonanten op, bijvoorbeeld na *-n* en *-l-*, maar wordt dan niet in het moderne lemma overgenomen (gezien het eerste voorbeeld ook bij flexie-uitgangen):

vnnl. veelderhande	VELERHANDE
nnl. twijnder	TWIJNER

Toelichting: moderne citaten hebben hier de *d*-loze vorm. Uitzonderingen zijn gelexicaliseerde lemmata als DIENDER, VILDER en vaktermen als STEUNDER; zij behouden de *-d-*.

9.3.8 Samenvall en splitsing

Voorbeeld 1:

horn heeft als kernbetekenis ‘dierenhoorn’ en heeft allerlei betekenissen rond ‘spits’. Het komt in het historische materiaal voor als *horn*, met rekking van de vocaal als *hoorn*, met sjwa-insertie (svarabahktivocaal) als *horen*. Er heeft zich samen met de vormwisseling betekenispecialisatie voorgedaan, waardoor het huidige Nederlands deels verschillende woorden heeft. HORN is het lemma voor ‘hoek, spits, punt’. Een argument voor HORN is ook het frequente HORNIK in oude bronnen. Daarnaast is HOORN het lemma voor ‘hoorn van dieren’, ‘blaasinstrument’, waarbij het lexicon ook een variant HOREN kent. In de samenstellingen en afleidingen idem.

Voorbeeld 2:

Vergelijkbaar zijn de doubletten waarbij er twee gelijkwaardige varianten ontstaan door invloed van de flexievormen (bijvoorbeeld een op basis van nominatief en een op basis van verbogen vormen). Die beide vormen vinden we in historisch materiaal terug. Als echter in het Standaardnederlands maar één van die vormen gebruikelijk is, zoals *geel* bij *geel-geluw* en *schaduw* bij *schade-schaduw* ‘silhouet, schim’, dan geldt alleen die moderne vorm als het lemma.

onl. skado	SCHADUW
mnl. scade, schaduwe	SCHADUW
nnl. schade, schaduw	SCHADUW

Bij samengestelde woorden wordt eveneens de moderne vorm als lemma toegekend. Zijn beide vormen in het moderne Nederlands nog bekend met betekenisdifferentiatie, dan worden die als twee verschillende woorden behandeld:

onl. nahtskado ‘nachtelijke schaduw’	NACHTSCHADUW
mnl. nachscade, nachscaduw ‘plant’	NACHTSCHADE
nnl. nachtschade, nachtschaduwe ‘plant’	NACHTSCHADE
nnl. nachtschaduw ‘nachtelijke schaduw’	NACHTSCHADUW

Bij 4 De lemmagrenzen

9.4.1 Enkele specifieke verbindingen: naamwoorden met een genitief

Flexievormen geven vaak uitsluitsel over de structuur:

(here van) tsgravensteen = des graven steen DE / GRAAF / STEEN

tsgravensteen (in Leiden) DE / GRAAF / STEEN

Het genitiefliedwoord kan hier niet op *gravensteen* betrekking hebben en hoort dus bij *graaf*. Dat kan wel in het onderstaande voorbeeld, waardoor *gravensteen* als geheel wordt opgevat.

(die here) tsgravensteens HET / GRAVENSTEEN

Zo ook in: (het) gravensteen (in Leiden) GRAVENSTEEN

9.4.2 Enkele specifieke typen: combinatie met een werkwoord

Of een verbinding met een werkwoord een samenkoppeling is, kan niet door een regel vastgesteld worden. Het criterium is de lexicalisering van de verbinding anno nu, met die aantekening dat hierbij in het hedendaags Nederlands relatief snel veranderingen optreden. Zo heeft Molex PLATLIGGEN, PLATLEGGEN, SCHOONWASSEN, SCHOONWRIJVEN, ZWARTKIJKEN, ZWARTMAKEN (hier, willekeurig, alleen voorbeelden van adjectief + werkwoord) Aangezien Molex niet pretendeert volledig te zijn in het weergeven van woorden met dezelfde structuur kunnen we uit deze voorbeelden niet per se afleiden of *schoon lebberen* of *schoonlebbere* geschreven moet worden. Als bestaande lexicale bronnen het desbetreffende woord ook niet hebben, gelden de gangbaarheid en de eenheidsbetekenis als criterium. Analogie, bijvoorbeeld vanwege de spelling van andere werkwoorden met *schoon*, kan ook een rol spelen:

hij dweilde de lei schoon SCHOONDWEILEN

Een probleem is dat veel samenkoppelingen in de taalhistorie nog verbindingen waren, wat kan leiden tot nogal onsystematische lemmatisering:

(iemand, iets) lief hebben LIEFHEBBEN (modern Nederlands)

(iemand, iets) liever (liefst) hebben LIEVER (LIEFST) / HEBBEN

(iemand) leed hebben LEED / HEBBEN

NB: de infinitief wordt niet noodzakelijk aaneengeschreven als er een aaneengeschreven deelwoord is: het lemma KAALGEGRAASD houdt niet per se in *kaalgrazen* en het lemma is BIERDRINKEND naast de meerwoordsexpressie BIER DRINKEN.

9.4.3 Enkele specifieke verbindingen: combinatie met een deelwoord

De norm bij de als adjectief gebruikte deelwoorden is ook steeds de lexicalisering van de verbinding anno nu en daarbij de huidige spellingsregels:

wel bespreekt WELBESPRAAKT

vast besloten VASTBESLOTEN

geschreven als één woord, net als BLAUWGEVERFD, WITGEKALKT, WITGELAKT, WITGEPELST. Verder wordt het deelwoord altijd aaneengeschreven als de infinitief aaneengeschreven wordt: SCHOONPOETSEN, dus SCHOONGEPOETST (andersom dus niet per se, zie ook 9.4.2).

De spellingregels kunnen hier niet eenduidig zijn en Molex bevat uiteraard niet alle mogelijke vormen, maar analogie en frequentie helpen om de schrijfwijze van dit type verbindingen te bepalen. Dat betekent dat *groengeverfd* aan elkaar geschreven wordt, maar dat er voor samenstellingen met minder gewone kleurnamen gekozen wordt voor los schrijven:

de <u>groen</u> <u>geverfde</u> deur	GROENGEVERFD
de <u>olijfgroen</u> <u>geverfde</u> deur	OLIJFGROEN / GEVERFD

Zie ook:

<u>hoogopgeleide</u> mensen	HOOGOPGELEID
<u>hoog opgegroeid</u> gewas	HOOG / OPGEGROEID
<u>slecht opgevoede</u> kinderen	SLECHT / OPGEVOED

Ook hier zorgt de diachronie voor inconsequenties. Bijvoorbeeld:

dit <u>schreit ten hemel</u>	SCHREIEN / TE / DE / HEMEL
dit is <u>ten hemel schreiend</u>	TENHEMELSCHREIEND

Dit laatste is een samenkoppeling.

Bij 5.2 Affixen

9.5.2.1 Afwijkingen

In bepaalde woorden kan er afgeweken worden van de gebruikelijke standaardisering en wordt een bijzondere vorm van het affix gebruikt:

Voorbeeld 4:

mnl. *begard*, *beggaert*, *bogaert*, nnl. *begard* ‘lekenbroeder’. Dit woord bevat het suffix *-aard*, later ook geëvolueerd tot *-erd*, maar heeft veelal de vorm *-ard* (mogelijk omdat het via het Latijn of Frans herontleend is). De meest recente vorm is *begard*; vandaar hebben alle genoemde woordvormen het lemma BEGARD.

9.5.2.2 Enkele specifieke suffixen 1: -e

-e is in de volgende gevallen een suffix en wordt dus gehandhaafd in het lemma:

a. bij (doorgaans uitheemse) afgeleide vrouwelijke persoonsnamen: SCHOLIERE, KAMENIERE, ASSISTENTE.

b. (uit \bar{t} , nu verouderd) bij abstracta afgeleid van adjectieven: HOGE ‘hoogte’, KOUDE, *KOU* ‘lage temperatuur’. Dus mnl., nnl. *evenhoge* znw. ‘bouwsel van gelijke hoogte’ (als verdedigingswerk) heeft het lemma EVENHOGE. (Hier zou het eventueel ook een gesubstantiveerd adjectief kunnen zijn.)

c. bij verouderde afleidingen voor mannelijke persoonsnamen.

mnl. herde	HERDE
mnl. schutte	SCHUTTE
mnl. wrechte	WRECHTE

In deze woorden is *-e* later vervangen door *-er*. Verder nog woorden als SCHINKE, KEMPE, ERVE, WINNE (zie De Vooy 1967: 203). Soms komen er *-e-loze* vormen voor, zoals in het WNT WIN(I); het lemma wordt toch WINNE.

NB 1: eveneens een suffix is het mnl. bijwoordelijke *-e* (onl. *-o*), maar dit wordt niet gelemmatiseerd, met uitzonderingen als *VERRE* (een bijwoordelijke variant van het adjectief *VER*), dat nog voorkomt in ‘van heinde en verre’; dit is dus een uitzondering op de regel van het handhaven van oude affixen! Deze uitgang is afgesleten en niet vervangen door een ander:

onl. armlīko	ARMELIJK
mnl. gevankelike	GEVANKELIJK

en vele andere woorden op mnl. *-like* / onl. *-līko* (die in het Nieuwnederlands dus soms samenvallen met adjectieven op *-lijk*).

NB 2: geen suffix, maar wel gehandhaafd in het lemma is de gelexicaliseerde uitgang van gesubstantiveerde adjectieven zoals *SCHONE* ‘mooie vrouw’, *BLINDE* ‘niet-ziende persoon’ en in (op adjectieven gebaseerde) vrouwelijke inwonersnamen als *LEIDSE*, *BERLIJNSE*.

Zie verder de voorbeelden onder *e*-apocope, 9.3.2.

9.5.2.3 Enkele specifieke suffixen 2: *-(e)de/-te*

Voor abstracta op *-(e)de/-te* (uit *-itha*) (zie De Vooy 1967: 215-216) wordt in de moderne lemmatisering één vorm gehanteerd: *-te*, als resultaat van een klankontwikkeling (In de historische woordenboeken zelf is dit verschillend gedaan).

Voorbeelden:

nl. hoogde	HOOGTE
mnl. hogede	HOOGTE
mnl. lengede	LENGTE
onl. diupitha	DIEPTE

Afleidingen van vergelijkbare adjectieven, zoals *oud*, *lief*, *vol*, *dik*, *oud*, *rond*, *koud*, *stil*, *zwaar*, worden systematisch gelemmatiseerd met *-e* of *-te*.

Soms staat een *-e*-abstractum naast een *-te*-abstractum en is er dus sprake van twee verschillende woorden: bij *lang*: *LANGE* en *LENGTE*, bij *lief*: *LIEVE* en *LIEFDE*.

Soms is er geen *-te*-variant: bij *koud*: alleen *KOUDE* met variant *KOU*, bij *gezond* alleen *GEZONDE*, bij *waard* alleen *WAARDE*, of er is géén *-e*-variant: bij *steil* alleen *STEILTE*.

In het volgende geval is er ook een aan het Duits ontleende vorm en zijn er dus drie woorden met eigen lemma: bij *gemeen*: *GEMENE*, *GEMEENTE* en *GEMEENDE*.

Soms valt de *-e* vorm samen met gesubstantiveerd adjectief op *-e* of komt ernaast voor, of is er alleen maar dat gesubstantiveerd adjectief (MNW *GROENE(II)*).

9.5.2.4 Enkele specifieke suffixen 3: feminiserende suffixen

Afgeleide vrouwelijke persoonsnamen krijgen altijd een eigen lemma en worden naar suffix ook altijd onderscheiden (dat gebeurde in de woordenboeklemma's niet altijd). Dus:

lererse	LERERSE (niet: LERARES)
---------	-------------------------

De volledige reeks suffixen is in standaardvorm: *-in*, *-es* (heel soms *-esse*), *-erse* (< *-eresse*), *-se* (< *-ske*, *-sche*), *-egge* (< *ige*, *-icghe*, *-ege* etc.), *-e*, *-nede*, *-ster*. Bij het onderscheid tussen *-se* en *-erse* moet soms naar de woordvorm worden gekeken. Er kunnen ook stapelsuffixen voorkomen (bijvoorbeeld *-steregge*).

9.5.2.5 Enkele specifieke suffixen 4: diminutiefsuffixen

Suffixen die geëvolueerd zijn naar het moderne *-tje* krijgen steeds die vorm (of een allomorf). Dit geldt voor *-kijn*, *-kin*, *-ken*, *-sken*, *-gien*, *-gen*, *-tjen*, etc. In diminutiefsuffixen die verdwenen zijn in het Standaardnederlands en die ook het element *-īn* bevatten, wordt dit als *-ijn* weergegeven: *-ijn*, *-sijn*, *-(e)lijn*.

Voorbeelden met *-je*:

mnl. knaepgen, knaapkens	KNAAPJE
mnl. paeslamken	PAASLAMMETJE

Voorbeelden met *-ijn*:

mnl. welpijn	WELPIJN
mnl. bergelyn	BERGELIJN
mnl. hoecsijn	HOEKSIJN

9.5.2.6 Leensuffixen en suffixen in leenwoorden

Hier wordt per woord omgezet, omdat woorden met deze suffixen vaak via verschillende wegen ontleend worden; de suffixen hebben geen vaste vorm. Soms heeft het ene lemma één vorm en een ander lemma een andere vorm.

Voorbeeld 1: fra. *-on* > nl. zowel *-on* als *-oen*.

Bastoen (ofr. *baston*) in VMNW, MNW, *baston* in WNT; aldaar ook *bastonneren* en *bastonnade*. Het lemma is dus BASTON.

Voorbeeld 2: fra. *-ir*: De oudere vorm is meestal *-ieren*, en later is deze in de regel *-eren*. In de lemma's hangt het af van hoe het nu (of in het WNT) voorkomt of het lemma *-ieren* of *-eren* krijgt.

vmnl. purgieren	PURGEREN
vmnl., mnl, nnl. purgeren	PURGEREN
vmnl. obligieren	OBLIGEREN
vmnl., mnl, nnl. obligeren	OBLIGEREN
mnl. faelgieren, faelgeren	FAILLEREN
nnl. tabagieren	TABAGIEREN
mnl., nnl. drapenieren	DRAPENIEREN
nnl. kamenieren	KAMENIEREN

De laatste twee woorden zijn afgeleid van het zelfstandig naamwoord op *-ier*.

Voorbeeld 3: Romaans/Oudfrans suffix *-oel* (zie VMNW *halsbergoel*, *stassoel*) komt ook via jonger Frans als *-ole*, *-ool* voor (MNW *citole*, WNT *kamizool*). Daarnaast is er ook een woord *karakol* dat via het Spaans is ontleend (met mv. op *-ols* en *-ollen*). Deze worden niet alle op *-ool* gelemmatiseerd maar per woord bekeken:

vmnl., mnl. halsbergoel	HALSBERGOEL
vmnl. wambesoel, mnl. wamsool	WAMSOEL
mnl. citole	CITOO (met e-apocope)
nnl. karakol, krakol, karkol	CARACOLE (modern Nederlands)

Voorbeeld 4: fr. *-ien* bij inwonersnamen; hier wordt in het moderne lemma het Franse suffix aangepast, analoog aan latere ontwikkelingen in het Nederlands.

vmnl, mnl. galilien
nml. parizijn
mnl. cahorsijn

GALILIJN
PARIZIJN (als soortnaam PARIZIJN)
CAHORSIJN (als soortnaam KAUWERSIJN)

waarbij CAHORSIJN verwijst naar de inwoner van Cahors en KAUWERSIJN naar ‘geldwisselaar, woekeraar’.

Bij 5.3 Vreemde woorden

9.5.3.1 Voorbeelden van aanpassingen

Voorbeeld 1, bij ofr. *vaissel(e)ment* ‘vaatwerk’:

VMNW *facelment*, MNW *facelment*, verwijslemma *vaselment*; WNT *vassel(le)ment*, verwijslemma *vaselment*: dit wordt één lemma: VASSELEMENT (soort grote gemene deler).

Voorbeeld 2, aanpassingen aan modern Nederlands:

Een oorspronkelijke, maar in het Nederlands weggevallen klinker of lettergreep wordt niet toegevoegd. Er komen dus aparte lemma’s: SELEREN naast ASSELEREN.

Als beide vormen van een woord gelexicaliseerd zijn en in de woordenboeken opgenomen, blijven dat afzonderlijke lemma’s:

mnl., nml. dangier (< fr. dangier)	DANGIER
mnl., nml. angier (< fr. dangier)	ANGIER

Toelichting bij *angier*: dit is een voorbeeld van metanalyse; *d-* is opgevat als lidwoord.

Per geval moet gekeken worden of het woord (of een deel ervan) moderne varianten kent: mnl. bascoert > BASKOERT, en niet BASCOURT (< ofr. *basse-court*), omdat het WNT ook nog het simplex KOERT kent.

Voorbeeld 3:

Een leenwoord kan de vorm hebben van het oorspronkelijke vreemde woord, maar later daarnaast meer of zelfs geheel vernederlandst zijn. Het gaat dan om verschillende woorden met elk een eigen lemma, zoals *fenegriek* naast *fenigrecum* (< lat. *foenum graecum* ‘Grieks hooi’).

vmnl. fenigrecum	FENIGRECUM
mnl. fenigreec, venigriec	FENEGRIEK
nml. fenegriek	FENEGRIEK

9.5.3.2 Verdwenen vreemde woorden

Verdwenen vreemde woorden volgen in principe ook de klankwijzigingen, zoals *e*-apocope, als er een vorm met zo’n klankwijziging aangetroffen is:

mnl. <i>daet</i> (< fr. date, lat. datum)	DAAT
Zo ook mnl., nml. <i>date</i>	DAAT

Als van een verdwenen leenwoord niet goed uit te maken is waarheen het geëvolueerd zou zijn, dan moeten we behoudend lemmatiseren en de historische vorm overnemen.

mnl. <i>seme</i> ‘barbeel’	SEME
mnl. <i>semeil</i> (onzekere betekenis)	SEMEIL
mnl. <i>bise</i> ‘hinde’ (< fr. biche)	BISE

9.5.3.3 Voorbeelden van meervoudige ontlening

Voorbeeld 1, meervoudige ontlening met minimaal klankverschil, en toch verschillende lemma's:

mnl., nnl. porie	PORIE
mnl. pore, porre	POOR

Pore < lat. *porus*. Waarschijnlijk is later het Latijnse meervoud *pori* als een enkelvoud opgevat, waarna er stapelmeervoud *poriën* kwam, en met rückbildung daarnaast een enkelvoud *porie*. Niet alle details van de etymologie worden uitgedrukt in het lemma. Zie *mole* – *molen* (met verschillende etymologieën!).

Voorbeeld 2:

onl. mola (< ofr. mole < lat. mola)	MOOL
onl. mulina (< mlat. molina, mulina)	MOLEN
vmnl. muele (< lat. molinae, oude ontl.)	MOOL
vmnl. muelin (< mlat. molina, jonge ontl.)	MOLEN
mnl. mole	MOOL
mnl. molene	MOLEN
nnl. molen	MOLEN

Samenstellingen volgen hetzelfde principe. WATERMOLEN krijgt steeds een lemma op *-en*, ook al komt het met en zonder *-n* voor als historisch lemma. MOLEGAREN bevat waarschijnlijk een ander eerste lid. Het EWN interpreteert *mole* overigens als resultaat van herinterpretatie van de buigingsuitgang.

Voorbeeld 3:

In enkele gevallen zijn de vormen niet goed uit elkaar te halen en wordt één lemma toegekend. Bijvoorbeeld bij de vormen van *glos*, *glosse*, *gloze* (gebaseerd op twee Latijnse vormvarianten die elkaar in de tijd afwisselden) en hun afleidingen en samenstellingen wordt het lemma GLOS-, waarbij de lemmavariant *GLOSSE* ook in het lexicon wordt opgenomen.

Voorbeeld 4: Ook als het om een vormvariant gaat, zonder betekenisverschil, kennen we verschillende lemma's toe:

mnl. epelence, epilenci, hipilentie	EPILENTIE
nnl. epilentie	EPILENTIE

maar:

nnl. epilepsie	EPILEPSIE
----------------	-----------

De vorm *epilepsie* heeft *epilentie* in de loop van de tijd vervangen (EWN hierover: “De vormen met *-ms-* en *-nc-* zijn waarschijnlijk overgenomen uit Oudfrans *epilence* [1265; Rey], dat via het vulgair Latijn is ontstaan uit hetzelfde Latijnse woord [*epilepsia*], en door de medische wetenschap daar later door is vervangen.”).

Literatuur

- Bree, C. van (2004). *Historische taalkunde (2nd edition)*. Amersfoort: Acco.
- De Grave, J.J.S. (1906). *De Franse woorden in het Nederlands*. Amsterdam.
- Depuydt, K., Tiberius, C. & Heylen, K. (to appear). 'Lexicography of the Dutch Language: from dictionaries to a lexicographical infrastructure.' In H. Nesi & P. Milin (Eds.), *International Encyclopedia of Language and Linguistics (3rd Edition)*.
- Haas, W. de & Trommelen, M. (1993). *Morfologisch Handboek van het Nederlands*. 's-Gravenhage: SDU Uitgeverij.
- Haga, T., Depuydt, K. & Does, J. de, m.m.v. Bonth, R. de, & Geirnaert, D. (2024). *Tagset voor Diachroon corpusmateriaal van het Nederlands (TDN)*. INT Working Papers 1 (Versie 2.0) (Instituut voor de Nederlandse Taal).
- Horst, J.M. van der (2008). *Geschiedenis van de Nederlandse syntaxis*. Leuven: Universitaire Pers.
- Karsdorp, F. B. (2011). *Achtergronden bij de morfologische module van GiGaNT*. Leiden, Instituut voor Nederlandse Lexicologie. Interne nota.
- Loey, A. van (1964). *Schönfelds Historische Grammatica van het Nederlands*. Zutphen: Thieme.
- Loey, A. van (1976). *Middel nederlandse spraakkunst. Deel II. Klankleer (7^{de} herziene druk)*. Groningen: Tjeenk Willink.
- Loey, A. van (1980). *Middel nederlandse spraakkunst. Deel I. Vormleer (9^{de} druk)*. Groningen: Tjeenk Willink.
- Mooijaart, M. (2010). 'The complete history? Dutch words in four historical dictionaries'. In J. Considine (Ed.), *Current projects in historical lexicography*. Newcastle: Cambridge Scholars Publishing, 83-98.
- Mooijaart, M. (2011). 'Van acht (507) tot zwangerschap (1851): historische lexicografie van boek tot internet'. In *Nederlandse Taalkunde* 16 (2) (Discussie Historische woordenboeken op internet), 141-152.
- Mooijaart, M. (2012). Het lemma in het GiGaNT-lexicon. *INL Working Papers - Taalbank Nederlands 4*. Leiden.

- Pijnenburg, W. J. J., van Leuvensteijn, J. A., & van der Horst, J. M. (1997). *Geschiedenis van de Nederlandse taal*. M. C. van den Toorn (Ed.). Amsterdam: Amsterdam University Press.
- Ruitenbergh, T. & Van pellicom, K. (2012). De morfosyntactische module van het GiGaNT-lexicon. *INL Working Papers - Taalbank Nederlands 3*. Leiden.
- Pijnenburg W. (1993). '100 jaar *achtig*. Over de wording van een samengesteld suffix', in: *ABäG 37*, 33–48.
- Ruitenbergh, T., Does, J. de & Depuydt, K. (2010). 'Developing GiGaNT, a lexical infrastructure covering 16 centuries'. In: *Proceedings of the XIV Euralex International Congress*. Leeuwarden: Fryske Akademy, 468-476.
- Ruitenbergh, T. en Van pellicom, K. i.s.m. Does, J. de en Depuydt, K. (2012). *De morfosyntactische component van het GiGaNT-lexicon*. Leiden, Instituut voor Nederlandse Lexicologie. Interne nota.
- Sijs, N. van der (2004). *Taal als mensenwerk. Het ontstaan van het ABN*. Den Haag: Sdu Uitgevers.
- Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Vooyoys, C. G. N. de, & Schönfeld, M. (1967). *Nederlandse spraakkunst*. Ongewijzigde zevende druk, van vierde druk 1957. Groningen: Wolters.
- Wal, M. van der & Bree, C. van (2008). *Geschiedenis van het Nederlands*. Houten: Het Spectrum.
- Willemys, R., & Daniëls, W. (Ed.) (2003). *Het verhaal van het Vlaams, de geschiedenis van het Nederlands in de Zuidelijke Nederlanden*. Antwerpen/Utrecht: Standaard/Het Spectrum.

Verder geraadpleegd:

Spelling:

Technische handleiding: <https://woordenlijst.org/th.php>

Woordenlijst Nederlandse Taal: [Woordenlijst](#)

Historische woordvormen en etymologie:

Historische woordenboeken van het Nederlands met inleidingen (ONW, VMNW, WNT, MNW): <https://gtb.ivdnt.org/>

Etymologische woordenboeken:

Sijs, N. van der (samensteller) (2010). *Etymologiebank*, op www.etymologiebank.nl.